

Development of novel WWP2 ubiquitin ligase inhibitors using structural and biochemical approaches

JESSICA ELLON WATT

A thesis submitted for the degree of Doctor of Philosophy at the University of East Anglia, School of Biological Sciences, September 2018

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived therefrom must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract

A fundamental modification called ubiquitination controls the fate of many proteins. The third enzyme in the ubiquitination pathway, the E3 ubiquitin ligase, has been implicated in a variety of cancers. WWP2 is a HECT E3 ligase associated with the progression of prostate cancer, alongside its closest relative WWP1. WWP2 selectively targets substrates via the interaction of WW domains with PPxY motifs. Ubiquitin is consequently added to the substrate using the catalytic cysteine in the HECT domain. The three isoforms of WWP2 with different WW domain compositions have shown selective preferences towards substrates of the TGF- β signalling pathway and PTEN, a negative regulator of AKT signalling. Understanding more about the WW domain interactions and the activity of the HECT domain can lead to the discovery and development of novel therapeutics.

The aim of this thesis is to find novel inhibitors of WWP2 by utilising different approaches. The initial approach was to structurally solve WWP2 WW domains that reside within different isoforms and investigate substrate-based peptide interactions. Using structural analysis by NMR spectroscopy, the solved structure of WW4 is reported here. NMR titrations were used for interaction analysis of WW1 and WW1-2 domains, with substrate-based peptides, to aid peptide-based inhibitor design and characterise the selectivity of the WW domains. The other strategy involved small molecule screening to find inhibitors of WWP2-FL activity. Results presented here identify potential inhibitors with IC₅₀ values below 3 μ M. Counter screening against other enzymes in the ubiquitination cascade was also carried out to investigate the selectivity of the hit compounds. A total of 5 final compounds were successfully identified as selective, making them attractive for further investigations. Overall, the data collected in this thesis can be used in further understanding WWP2 interactions and for structure-based/hit-to-lead development for novel inhibitory molecules as potential prostate cancer therapeutics.

Table of Contents

Abstract.....	ii
Table of Contents	iii
List of Figures.....	vi
List of Tables	ix
List of Abbreviations.....	xi
Acknowledgements.....	xiv
1. Introduction.....	1
1.1. Ubiquitination	3
1.2. E1-activating and E2-conjugating enzymes	6
1.3. E3 ligases.....	8
1.3.1. RING E3 ligases.....	8
1.3.2. HECT E3 Ligases.....	11
1.4. Nedd4 family of HECT E3 ligases	15
1.4.1. WWP2 and its isoforms.....	20
1.4.2. WWP2, signaling pathways and their role in prostate cancer	22
1.5. TGF- β signalling	26
1.5.1. TGF- β signaling in prostate cancer	30
1.6. PTEN and PI3K/AKT signaling	32
1.6.1. PTEN in prostate cancer.....	35
1.7. Inhibitors of ubiquitination	36
1.8. Thesis Aims.....	38
2. Material and Methods	39
2.1. DNA-based procedures	40
2.1.1. Plasmids and primers	40
2.1.2. Phusion PCR.....	43
2.1.3. DNA gel extraction	43
2.1.4. In-fusion.....	43
2.1.5. Restriction digest.....	44
2.1.6. QuikChange mutagenesis.....	44
2.1.7. Heat shock transformation	45
2.1.8. Plasmid purification.....	45
2.2. Protein-based procedures	46
2.2.1. Recombinant protein expression - unlabelled	46

2.2.2. Recombinant protein expression - labelled	47
2.2.3. Cell lysate preparation	48
2.2.4. Protein purification methods	48
2.2.5. GST-pull down assay	50
2.2.6. SDS-PAGE.....	50
2.2.7. Western blotting	52
2.2.8. Glutathione plate coating	53
2.2.9. In vitro auto-ubiquitination assays	53
2.2.10. Statistical analysis.....	55
2.3. Cell-based procedures.....	56
2.3.1. Cell line preparation	56
2.3.2. Cell maintenance	56
2.3.3. 6-24 well plate preparation.....	56
2.3.4. Scratch assays.....	57
2.4. NMR-based procedures.....	58
2.4.1. NMR sample preparation	58
2.4.2. NMR spectroscopy experiments.....	58
2.4.3. NMR titrations	59
2.4.4. Processing and analysis of spectra	62
2.4.5. Structural calculations.....	63
2.4.6. Structural validation.....	64
3. Structural Characterisation of WW4	65
3.1. Introduction	66
3.2. Results	72
3.2.1. HSQC analysis for sample comparison of double labelled WW4	72
3.2.2. Assignment of GB1-WW4 side chains.....	73
3.2.3. ARIA: automated peak assignment, structural calculations and water refinement.....	73
3.2.4. Structure Validation.....	78
3.3. Discussion.....	87
4. WW domain interaction analysis	91
4.1. Introduction	92
4.2. Results	98
4.2.1. WWP2-PTEN interaction analysis	98
4.2.2. Smad peptide generation.....	104
4.2.3. GB1-WW1 backbone assignment.....	107
4.2.4. Titration analysis of WW1 interactions with Smad protein-based peptides	

4.2.5. Partial assignment of ¹⁵ N-GB1-WW1-2	116
4.2.6. Titration analysis of ¹⁵ N-GB1-WW1-2 tandem domain with Smad-based peptides	118
4.3. Discussion.....	129
5. High-throughput screening of compounds for WWP2 inhibitor discovery	137
5.1. Introduction	138
5.2. Results	141
5.2.1. Auto-ubiquitination biochemical assay optimisation	141
5.2.2. Initial screening of compound libraries	146
5.2.3. Secondary screening of potential hits.....	148
5.2.4. Dose-dependent analysis of hit compounds.....	149
5.2.5. E2 and E1 counter screens.....	152
5.2.6. Nedd4 and WWP1 counter screens	156
5.2.7. Compound inhibition of substrate ubiquitination.....	160
5.2.8. Effect of compounds on advanced stage prostate cancer cell growth and migration.....	162
5.3. Discussion.....	175
6. Discussion.....	184
6.1. Introduction	185
6.2. Structure-based inhibitor discovery	185
6.3. Small-molecule based inhibitor discovery	198
6.4. Future work	206
References	207

List of Figures

Figure 1.1.1 - Step by step pathway diagram of ubiquitination	5
Figure 1.3.1 - Schematic diagrams of the RING domain and example ligases.....	10
Figure 1.3.2 - HECT domain structures for E6AP and Nedd4L when bound to E2/E2~Ub	12
Figure 1.4.1 - Members and domain arrangements of the Nedd4 E3 ligase family ...	15
Figure 1.4.2 - Comparison of the XP residues of Smad 7 bound WW domains in Nedd4 family members	19
Figure 1.4.3 - Diagrams defining the regions present in WWP2 isoforms.....	21
Figure 1.6.1 - Schematic representation of canonical TGF- β signaling	27
Figure 1.6.2 - Schematic representations of Smad protein domains	28
Figure 1.7.1 - Schematic representation of PI3K/AKT signaling activation	33
Figure 3.1.1 - Schematic representations of a general 2D NMR experiment.....	69
Figure 3.2.1 - Comparison of double-labelled ^{13}C - ^{15}N -GB1-WW4 samples by HSQC.	73
Figure 3.2.2 - Initial partially refined GB1 and WW4 ensemble alignment from CCPN grid run 1.	75
Figure 3.2.3 - Final refined ensemble for GB1-WW4 from the WeNMR ARIA 1.2 HJ webportal calculations.	77
Figure 3.2.4 - Ensemble comparison of GB1 in GB1-WW4 with published 3GB1.	78
Figure 3.2.5 - MolProbity Ramachandran plots representing the analysis of 20 water refined models of GB1-WW4.	81
Figure 3.2.6 - wwPDB validation result summaries for refined GB1-WW4 ensemble.	84
Figure 4.3.1 - WW4 structure showing XP groove and other conserved residues.....	89
Figure 4.1.1 - Sequence comparison of WWP2 WW domains.	94
Figure 4.2.1 - WWP2 isoform interaction analysis shows binding differences.	99
Figure 4.2.2 - Pull-down analysis of PTEN deletion construct interactions with WWP2 isoforms N and C.....	101
Figure 4.2.3 - GST-pull-down analysis of tandem WW domains with PTEN constructs.	103
Figure 4.2.4 - His-SUMO-Smad 2 and -Smad 3 peptide construct purification by Ni- NTA affinity purification.....	105
Figure 4.2.5 - Digestion analysis of Smad 2 and Smad 3 peptide constructs.....	106
Figure 4.2.6 - Synthetic peptide sequences for Smad 2, 3 and 7 used in titration analysis.	107

Figure 4.2.7 - Assigned peaks of ^{15}N -GB1-WW1.	108
Figure 4.2.8 - $[\text{}^1\text{H}-^{15}\text{N}]$ -HSQC comparison of a ^{13}C - ^{15}N -GB1-WW1 sample and ^{15}N -GB1-WW1 sample.....	110
Figure 4.2.9 - Titration of GB1-WW1 with Smad 2 peptide.....	111
Figure 4.2.10 - ^{15}N -GB1-WW1 titration with increasing Smad 3 peptide additions. .	113
Figure 4.2.11 - Titration points 1:0 to 1:2 for ^{15}N -GB1-WW1: Smad 3.	114
Figure 4.2.12 - Titration HSQCs of ^{15}N -GB1-WW1 with increasing concentrations of Smad 7 peptide.	115
Figure 4.2.13 - $[\text{}^1\text{H}-^{15}\text{N}]$ -HSQC with partial backbone assignment for ^{15}N -GB1-WW1-2.	117
Figure 4.2.14 - GB1-WW1, -WW2 and -WW1-2 HSQC spectra overlaid.....	118
Figure 4.2.15 - Titration analysis of ^{15}N -GB1-WW1-2 with Smad 2 peptide shows protein denaturation.....	120
Figure 4.2.16 - $[\text{}^1\text{H}-^{15}\text{N}]$ -HSQCs from 1:0 to 1:4 for ^{15}N -GB1-WW1-2 titration with Smad 2 peptide before denaturation.....	121
Figure 4.2.17 - ^{15}N -GB1-WW1-2 titration with increasing Smad 3 peptide concentration leads to changes in peak chemical shifts.....	123
Figure 4.2.18 - Increasing Smad 3 peptide concentration leads to chemical shift perturbation of ^{15}N -GB1-WW1-2.....	124
Figure 4.2.19 - $[\text{}^1\text{H}-^{15}\text{N}]$ -HSQC spectra of increasing Smad 7 peptide concentration against ^{15}N -GB1-WW1-2.	127
Figure 4.2.20 - Comparison of $[\text{}^1\text{H}-^{15}\text{N}]$ -HSQC spectra of WW1-2: Smad7 at 1:11 with WW1-2: Smad 2 1:4 and WW1-2: Smad 3 1:7.5.....	128
Figure 4.3.1 - Jpred 4 simplified representation of WW1 and WW1-2 predicted structures.	135
Figure 5.2.1 - GST-WWP2-C auto-ubiquitination and inhibition test.....	142
Figure 5.2.2 - GST-WWP2-C plate assay optimisation.	144
Figure 5.2.3 - GST-WWP2-FL plate assay optimisation.....	145
Figure 5.2.4 - Compound screens using GST-WWP2-FL auto-ubiquitination assay.	147
Figure 5.2.5 - Triplicate assays showing best potential hits from initial screen.....	149
Figure 5.2.6 - Dose-dependent assays on 15 hits.....	150
Figure 5.2.7 - His-UbcH7 and GST-Uba1 counter assay optimisations.....	153
Figure 5.2.8 - Compound analysis against His-UbcH7 and GST-Uba1.	154
Figure 5.2.9 - His-Uba1-ubiquitin charging assay.	156
Figure 5.2.10 - HECT E3 ligase auto-ubiquitination counter screen optimisation. ...	157

Figure 5.2.11 - GST-Nedd4 and His-WWP1 compound assays.....	159
Figure 5.2.12 - GST-PTEN ubiquitination inhibition by compounds.....	161
Figure 5.2.13 - PC-3 scratch assays with diversity set compounds in the absence of TGF- β stimulation.....	163
Figure 5.2.14 - PC-3 scratch assays under TGF- β positive conditions.....	165
Figure 5.2.15 - Inhibition of PC-3 wound healing with natural products without TGF- β stimulation.....	167
Figure 5.2.16 - Inhibition of PC-3 wound healing with natural products under TGF- β positive conditions.....	168
Figure 5.2.17 - DU145 scratch assays under TGF- β negative conditions.....	170
Figure 5.2.18 - Diversity set compound DU145 scratch assays in the presence of TGF- β	171
Figure 5.2.19 - DU145 scratch assay analysis of natural products 6 and 17 under TGF- β negative conditions.....	173
Figure 5.2.20 - Natural product DU145 scratch assay analysis in the presence of TGF- β	174
Figure 5.3.1 - Structural modelling of the putative binding site in HECT domain of WWP2.....	178
Figure 6.2.1 - Structural comparison of WWP2 WW2, WW4 and HECW1 WW2 domains.....	186
Figure 6.2.2 - XP pocket comparison of WWP2 WW4 models with the second WW domains of WWP2 and HECW1.....	188
Figure 6.2.3 - Comparison of residues outside of the XP pocket involved in substrate interactions.....	189
Figure 6.2.5 - Comparison of the hydrophobic underside of WWP2 WW4 with WWP2 WW2 and HECW1 WW2.....	190

List of Tables

Table 2.1.1 - Summary of plasmid names and sources.	40
Table 2.1.2 - Summary of plasmids cloned during this work and the method used. ...	41
Table 2.1.3 - Summary of primers used in plasmid cloning and mutagenesis.	42
Table 2.1.4 - Components used for both small and large-scale phusion reactions (rxn).	43
Table 2.1.5 - Cycling conditions used for phusion reactions.....	43
Table 2.1.6 - Components used for a single reaction (rxn) mix.	44
Table 2.1.7 - Restriction digest reaction components per 20 μ L reaction (rxn).	44
Table 2.1.8 - Component volumes used for a single mutagenesis reaction (rxn).	45
Table 2.2.1 - Protein expression conditions used for different plasmids.	47
Table 2.2.2 - Gel components and volumes per 2 gels at 1.5 mm thickness.	51
Table 2.2.3 - Components and volumes used for making 2 tricine gels at 0.75 mm thickness.	51
Table 2.2.4 - Summary of antibodies, dilutions and source used.	53
Table 2.3.1 - Mammalian cell lines and culture media used in this research.	56
Table 2.4.1 - NMR sample preparation components and volumes.	58
Table 2.4.2 - Acquisition parameters for GB1-WW1 NMR experiments.	59
Table 2.4.3 - Acquisition parameters for GB1-WW4 NMR experiments.	59
Table 2.4.4 - Residue sequences for the recombinant peptides and synthetic peptides used in expressed-peptide generation and NMR titration analyses, respectively.	60
Table 2.4.5 - Conditions used during GB1-WW1 titration with synthetic Smad 2 peptide.	60
Table 2.4.6 - Peptide additions and concentrations throughout the GB1-WW1 titration using synthetic Smad 3 peptide.	60
Table 2.4.7 - Conditions used during the GB1-WW1 titration with synthetic Smad 7. 61	
Table 2.4.8 - GB1-WW1-2 titration conditions using synthetic Smad 2 peptide.	61
Table 2.4.9 - Concentrations used, and peptide additions during GB1-WW1-2 titration with synthetic Smad 3 peptide.	62
Table 2.4.10 - Smad 7 peptide additions and concentrations in GB1-WW1-2 titration.	62
Table 2.4.11 - Parameter settings used for the WeNMR ARIA 1.2 HJ webportal structural calculations.	64

Table 3.2.2 - Summary of PSVS validation of GB1-WW4 and individual GB1 and WW4 core regions.....	80
Table 3.2.3 - Summary of the criteria and validation results carried out by wwPDB on GB1-WW4 ensemble.....	85
Table 5.2.1 - Diversity set compound NSC codes, structures and IC50 values derived from dose-dependent analysis.....	151
Table 5.2.2 - Natural products NSC codes, structures and IC50 values derived from dose-dependent analysis.....	152

List of Abbreviations

Å	- Angstrom (10^{-10}m)
AAD	- Active adenylation domain
AKT	- Protein Kinase B
AMP	- Adenosine monophosphate
APC	- Anaphase-promoting complex
APS	- Ammonium persulphate
AR	- Androgen receptor
ATP	- Adenosine triphosphate
BARD1	- BRCA1 associated RING domain 1
BMP	- Bone Morphogenic protein
BMRB	- Biological Magnetic Resonance Bank
BRCA1	- Breast cancer susceptibility gene 1
C2 domain	- Calcium binding domain
CBL	- E3 ubiquitin ligase Cbl (Casitas B-lineage lymphoma proto-oncogene)
CBP	- CREB binding protein
CDK	- Cyclin-dependent kinase
CK2	- Casein kinase 2
Co-smad	- Common mediator/co-factor Smad
D ₂ O	- Deuterium oxide
DMEM	- Dulbecco's Modified Eagle's Medium
DMSO	- Dimethyl sulfoxide
DNA	- Deoxyribose nucleic acid
DPBS	- Dulbecco's Phosphate-buffered Saline
DSS	- 4, 4-dimethyl-4-silapentane-1-sulfonic acid
DTT	- Dithioltheritol
E6AP	- E6-associated protein
EDTA	- Ethylene diamine tetra-acetic acid
EGFR	- Epidermal growth factor receptor
EMT	- Epithelial to mesenchymal transition
ENaC	- Epithelial Sodium Channel
FAT10	- Ubiquitin D
FCS	- Foetal calf serum
FID	- Free induction decay
FL	- Full length

GSK	- Glycogen synthase kinase
GST	- Glutathione S transferase
H2A/B	- Histone 2A/2B
HECT E3	- Homologous to E6AP Carboxy terminus E3
HECW	- HECT, C2 and WW domain containing E3
HERC E3	- HECT and RLD domains containing E3
HSQC	- Heteronuclear single quantum coherence
IAD	- Inactive adenylation domain
IKK	- I κ B kinase
INEPT	- Insensitive nuclei enhanced by polarisation transfer
IPTG	- Isopropyl thiogalactoside
I-Smad	- Inhibitory Smad
ITC	- Isothermal calorimetry
Itch	- Itchy E3 ubiquitin ligase
kDa	- kilo Daltons
LB	- Lysogeny Broth
Mdm2	- Mouse double minute 2
MEM	- Minimal essential medium
MH1	- Mad homology 1
MH2	- Mad homology 2
MWCO	- Molecular weight cut off
NDFIP	- Nedd4 family interacting protein
Nedd4	- Neural precursor cell-expressed developmental downregulated gene 4
NF- κ B	- Nuclear factor κ B protein complex
NMR	- Nuclear magnetic resonance
NOESY	- Nuclear overhauser effect spectroscopy
OD	- Optical density
PAGE	- Polyacrylamide Gel Electrophoresis
PBS(T)	- Phosphate buffered saline (+Tween)
PCa	- Prostate cancer
PCR	- Polymerase chain reaction
PDB	- Protein Data Bank
PDK1	- Phosphoinositide dependent kinase 1
P/S	- Penicillin/Streptomycin
PI3K	- Phosphoinositide 3Mkinase
PIP ₂	- Phosphatidylinositol 4,5-bisphosphate

PIP ₃	- Phosphatidylinositol 3,4,5-trisphosphate
PKC	- Protein kinase C
ppm	- Parts per million
PSA	- Prostate specific antigen
PSVS	- Protein structure validation suite
PTEN	- Phosphatase and Tensin Homolog
RBR	- RING-between-RING
RCC1	- Regulator of chromatin condensation 1
RMSD	- Root mean square deviation
RNF4	- RING finger protein 4
rpm	- Revolutions per minute
RPMI 1640	- Roswell Park Memorial Institute 1640 medium
R-Smad	- Regulatory Smad
Rsp5	- Reverses SPT-phenotype protein 5 E3 ligase
RING E3	- RING finger-containing E3
SCF	- SCF ubiquitin ligase complex protein SKP1b
SDS	- Sodium Dodecyl Sulphate
SH2	- Src homology domain 2
SMURF	- Smad ubiquitin regulatory factor
Src	- Proto-oncogene tyrosine protein kinase
SSMPB	- Sulpho-succinimidyl 4-(p-maleimidophenyl) butyrate
TAE Buffer	- Tris acetate EDTA Buffer
TEMED	- Tetramethylethylenediamine
TFA	- Trifluoroacetic acid
TGF β	- Transforming growth factor β
TGF-βR	- Transforming growth factor β receptor
TOCSY	- Total correlation spectroscopy
TRAF6	- TNF receptor-associated factor 6
U-box	- UFD2 homology
Ub	- Ubiquitin
Uba1/Ube1	- Ubiquitin-activating enzyme E1
Ubc	- ubiquitin-conjugating
Ulp-1	- Ubiquitin-like protease 1
WW domain	- tryptophan-tryptophan recognition domain
WWP1	- WW domain containing E3 ubiquitin protein ligase 1
WWP2	- WW domain containing E3 ubiquitin protein ligase 2

Acknowledgements

Firstly, I would like to thank my supervisor, Dr Andrew Chantry, for all his help, guidance and patience throughout both the 3 years of research and the writing up period. I greatly appreciate how much his experience and positive thinking has pushed me onwards and impacted my scientific abilities for the better. I would also like to thank Dr Tharin Blumenschein for her wealth of knowledge in NMR and her support during this project. Her expertise and ability to break down the complex nature of NMR for a biologist to understand is something that I will not forget. I am pleased to have had the opportunity to work with both Andrew and Tharin and am grateful for their constructive criticism throughout my PhD. I'd also like to acknowledge the support Dr Charles Brearley has provided in overall progress reviewing and providing access to essential equipment during this time; as well as the Angulo lab for our collaboration in WWP2 inhibitor studies, and Dr Paul Thomas for his help regarding microscopy.

I'd like to thank both the past (Lloyd Wahl and Tiffany Yim) and present (Danielle Bourcier and Greg Hughes) members of the research group for putting up with my random mutterings in the lab, long discussions and chats over coffee. I appreciate all the fun times and reality checks, and for putting up with my (sometimes) ridiculous questions.

Lastly, I want to give special thanks to my family and friends outside of the lab who have provided me with all the support I could have asked for when things became challenging. I am thankful to my friends for all the laughs and motivation. Also, to my parents who never questioned my capabilities, continually supported me and provided me with wonderful food while I was writing up. Thank you to everyone mentioned here as I wouldn't be where I am now without your expertise, motivation and support.

1. Introduction

Proteins are one of the basic building blocks of cells. They have both structural and functional roles in many cellular processes which can vary across cell types. Regulation of proteins is a requirement for maintaining the correct environment for cells to function both internally and externally. Many mechanisms are employed in the regulation of protein activity from transcriptional regulation at the DNA level, to post-translational regulation at the protein level. This affects the localisation, activity and even the turnover of proteins. One such method of post-translational modification, which controls protein activity and ultimately the fate of the cell, is ubiquitination. This process can result in both proteolytic and non-proteolytic events downstream in signalling pathways. The addition of ubiquitin to the final substrate is carried out by the E3 ligase. The HECT E3 ligase called WWP2 (WW domain containing protein 2) is a member of the Nedd4 E3 ligase family. WWP2 acts by ligating ubiquitin monomers and polymers to substrates via intrinsic catalytic activity. This labels the target substrates for degradation by the proteasome, although other consequences may occur via WWP2 activity that have yet to be uncovered. Dysfunction and aberrant regulation of this ligase has been associated with disease. One of the most common diseases today is cancer which can develop either by genetic mutation or dysregulation at the protein level changing the cellular environment. This leads to cellular transformation and damaging effects in the body. The overexpression of WWP2 increases the targeting of tumour suppressor proteins which protect the cells from transformation. Known substrates of WWP2 are Smad proteins in the TGF- β signalling pathway, and PTEN, a negative regulator of PI3K/AKT signalling. Understanding more about how this ligase interacts and targets substrates in both the TGF- β and PI3K/AKT pathways will be explored further in this thesis with regards to prostate cancer development, which has been linked to WWP2 overexpression. This thesis aims to uncover the structural basis of the WW domains which interact with substrates as a method for progressing the design of short peptides that can potentially interfere with these protein-protein contacts. Another approach will also be explored using small molecules in the discovery of compounds which can inhibit WWP2 activity. This will be taken one step further to begin looking at the potential effects these compounds could have as possible treatments of advanced prostate cancer.

1.1. Ubiquitination

Many eukaryotic processes are regulated using modifications at the protein level. One such modification is ubiquitination, initially identified by the work of Ciechanover, Hershko and Rose (Ciechanover *et al.*, 1978; Hershko *et al.*, 1979; Hershko *et al.*, 1980) to play a role in protein stability and turnover. Ubiquitin addition has later been found to regulate other cellular processes as well such as signal transduction, transcriptional control and DNA repair (Boase and Kumar, 2015; Chantry, 2011). Ubiquitin (Ub) is a 76-amino acid protein which is highly conserved across eukaryotes, reflected in the sequence difference of only 3 amino acids between yeast and humans, supporting its function in cellular survival. Ubiquitin is covalently attached to the target substrate by three specialised enzymes: E1-activating, E2-conjugating and E3 ligase enzymes. Initially ubiquitin is attached to E1-activating enzyme in an ATP dependent manner forming E1~Ub thioester (Pickart and Eddins, 2004). The charged ubiquitin is then transferred to E2-conjugating enzyme (Berndsen and Wolberger, 2014), which consequently interacts with substrate bound E3 ubiquitin ligase, enabling the formation of an isopeptide bond between the ubiquitin molecule and a substrate lysine residue (Chantry, 2011; Kamadurai *et al.*, 2013; Soond *et al.*, 2013). Once complete, the E2 enzyme and substrate are released. This cascade can result in mono-, multi-mono- and poly-ubiquitination, each with specific biological functions related to the substrate. A simplified representation of this cascade can be seen in Figure 1.1.1.

Substrates can undergo different types of ubiquitination. The specific E2 and E3 enzymes involved in the cascade, timing, localisation and cellular context in addition to chain type and length can directly determine the fate of the substrate protein (Kirkin and Dikic, 2007; Nathan *et al.*, 2013). The presence of seven lysine (K) residues in ubiquitin at positions K6, K11, K27, K29, K33, K48 and K63 enables seven different inter-ubiquitin chain types (Schlesinger *et al.*, 1975; Komander and Rape, 2012). Additionally, evidence has been found for N-terminal linear ubiquitin chain formation creating a total of eight different linkage-types for the formation of poly-ubiquitin chains (Komander and Rape, 2012; Rieser *et al.*, 2013). The simplest type is mono-ubiquitination with the addition of a single ubiquitin monomer. This can also occur more than once across the target substrate referred to as multi-mono-ubiquitination. The role of mono-ubiquitination is not as well established as that of poly-ubiquitin chains, however Kaiser and colleagues (2011) discovered that the overall levels of mono-ubiquitination were higher than poly-ubiquitination supporting this modification type as an important regulator of cellular functions. This is also observed in the role of

mono-ubiquitination in non-proteolytic events (Nakagawa and Nakayama, 2015). For example, mono-ubiquitination of the histone protein H2B is required for transcriptional elongation, whereas the same modification to another histone protein H2A results in transcriptional repression by blocking H3K4 methylation (Nakagawa *et al.*, 2008). Another example is the tightly controlled localisation of the tumour suppressor protein p53, in which mono-ubiquitination results in conformational changes that expose the nuclear export signal inducing exclusion of p53 to the cytoplasm (Kruse and Gu, 2009). As for poly-ubiquitination, the length and type of chain seems to determine whether proteolytic or non-proteolytic events follow. K48 chains are the main chain type to target proteins for degradation by the 26S proteasome (Thrower *et al.*, 2000). A chain with a minimum of four ubiquitin molecules is required for sufficient proteasomal recognition by 19S subunit receptors leading to protein unfolding. Ubiquitin molecules are then recycled by deubiquitination (DUB) enzymes also associated with the 19S complex and the protein is then degraded. Complicating our understanding of the functions of ubiquitination types in cellular processes, however, is the emerging evidence of mono-ubiquitination also resulting in proteasomal degradation. This backs up the likelihood that it is more than just the ubiquitination type that determines proteolytic consequences (Livneh *et al.*, 2017). Support for the complexity of target substrate fate towards degradation also comes from evidence showing K63-Ub chains, historically said to mediate non-proteolytic activity, can trigger the internalisation of membrane proteins such as the receptor tyrosine kinase EGFR (epidermal growth factor receptor) for lysosomal degradation. Cytokine signal transduction is also regulated by K63 poly-ub chains, as observed for TRAF6-Ubc13-mediated K63 polyubiquitination resulting in the activation of IKK in the NF- κ B pathway (Deng *et al.*, 2000). Furthermore, K63-linked chains have been shown to stabilise histones at regions of DNA damage for the recruitment of DNA repair enzymes (Komander and Rape, 2012). Collectively, this shows the diverse and complex nature of ubiquitination and shows the requirement for further understanding of the mechanisms and the substrates that undergo this modification, both in normal and aberrant cellular conditions.

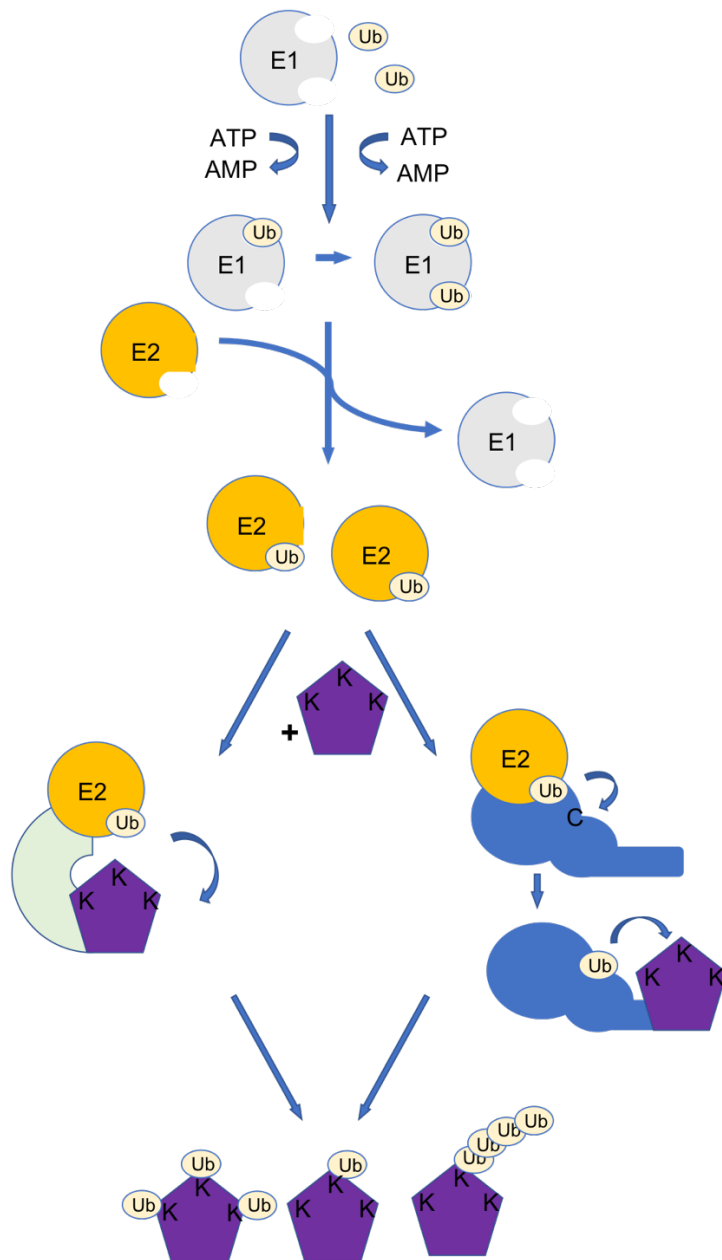


Figure 1.1.1 - Step by step pathway diagram of ubiquitination

Ubiquitin (Ub) is activated by the E1-activating enzyme (E1) and hydrolysis of ATP (adenosine triphosphate) to AMP (adenosine monophosphate) and pyrophosphate. Ubiquitin is then transferred to E2-conjugating enzyme (E2). Ubiquitin is then transferred in either a one-step (left pathway) or two-step reaction (right pathway) to the substrate via interaction of E2~Ub with an E3 ligase which also binds the substrate. Purple hexagon represents the substrate; green semi-circle denotes a RING E3 ligase; blue shape symbolises a HECT E3 ligase. C and K indicate the catalytic cysteine residue and target lysine residue, respectively. Mono-ubiquitination (bottom centre), multi-mono-ubiquitination (bottom left) and poly-ubiquitination (bottom right) are also shown.

1.2. E1-activating and E2-conjugating enzymes

As mentioned, there are three enzymes in the pathway that are required to ligate ubiquitin to specific targets. To date, only two E1 enzymes that interact with ubiquitin have been discovered (Schulman and Harper, 2009), each presenting slightly different roles in protein regulation. Uba1 (also known as UBE1), identified in 1979 (Hershko *et al.*, 1979), interacts directly with ATP and the ubiquitin molecule is adenylated before transfer to the thiol site on the E1 enzyme. Uba1 in humans is homologous to Uba1 in yeast which is composed of two adenylation domains, where one is active (AAD) and binds ATP and ubiquitin whilst the other is inactive (IAD) (Lee and Schindelin, 2008; Schafer *et al.*, 2014). Initially, ATP is hydrolysed, and the ubiquitin C-terminus adenylated. This then reacts with the catalytic cysteine residue to form a thioester bond between the ubiquitin carboxy terminal and the Uba1 cysteine side chain (Schulman and Harper, 2009). A second ubiquitin molecule then attaches to the initial adenylation site after the first has been attached to the catalytic cysteine residue (Haas *et al.*, 1982; Schafer *et al.*, 2014). This intermediate adenylated form is then maintained until transfer to the thiol site can occur. Purification of Uba1 led to the isolation of 2 isoforms (Cook and Chock, 1992) called Uba1a and Uba1b, generated from a secondary initiation codon in the gene locus. Each isoform has shown different localisation patterns. Uba1a localises to the nucleus whereas Uba1b, which lacks the first 40 amino acids where a nuclear localisation signal resides, mainly exists in the cytoplasm (Stephen *et al.*, 1997). This enables Uba1a interaction with specific E2 enzymes, which also localise to the nucleus, facilitating a nuclear ubiquitin-proteasome system for secondary messengers within the nucleus. As for Uba1b, this isoform acts outside of this organelle in the cytoplasmic proteasomal system. The other E1-activating enzyme, Uba6, is capable of activating both ubiquitin and the ubiquitin-like modifier FAT10. The ability of both of these E1 enzymes to bind to ubiquitin was thought to confer redundancy, however analysis of the E2 binding partners indicated this was only partially true. Knockout of Uba6 was embryonic lethal in mice via a defect in ubiquitin processing rather than FAT10 (Groettrup *et al.*, 2008), suggesting a fundamental Uba6 function in cellular processes involving ubiquitin that works alongside Uba1 isoforms. The implication this second E1-activating enzyme has on cellular regulation, and any potential overlap with Uba1 isoform activity, should be investigated to develop further understanding of the mechanisms and functions of ubiquitination, given the increasing evidence of protein turnover in disease development.

In contrast to the E1-activating enzymes, ~40 E2-conjugating enzymes exist in humans which further increases the selectivity of ubiquitination towards particular E3 ligases and substrate types (Stewart *et al.*, 2016). E2-conjugating enzymes have a central ubiquitin-conjugating (UBC) domain that contains all the essential binding sites and the catalytic cysteine residue required for function. A number of E2s also have additional regions, often disordered in nature. These extensions are thought to act as regulators of the E2 in aspects such as the intrinsic cysteine activity and ubiquitin chain generation, E3 interactions or even substrate selection (Pickart, 2001; Stewart *et al.*, 2016; Wenzel *et al.*, 2011). The noncovalent interaction of activated E1~Ub complex with E2 enables the transfer of ubiquitin to the E2 cysteine via a transthioesterification reaction. The activity of an E2 enzyme can be limited to a particular E3 ligase or a few, for example, some will only react with the catalytic cysteine of a HECT E3 ligase, but not a lysine residue as presented by RING ligases via the substrate molecule (Stewart *et al.*, 2016). This selectivity in the E2 activation mechanism refines the activities of the ubiquitination cascade to particular E3 ligases and their substrates. In turn, only the selective E3 ligases which interact with the given E2 enzyme are able to transfer the ubiquitin molecule to their substrate. This ultimately supports the role of E2-conjugating enzymes as regulators of E3 ligase activity and subsequent substrate modification consequences. A further regulatory mechanism of E2-conjugating enzymes is their inability to interact with both E1 and E3 enzymes at the same time due to an overlap in interaction sites (Wenzel *et al.*, 2011). This means that the E3 ligase has to be released before E2 can be charged with ubiquitin again and continue ubiquitin addition. While this may limit the processivity of the ubiquitination cascade, it provides a further stage of regulation to the pathway.

1.3. E3 ligases

The final step in the ubiquitination pathway is controlled by the activity of the E3 ubiquitin ligase. These enzymes are able to interact with both the E2 conjugating enzyme and the substrate, and therefore determine the target specificity of the activated pathway. Over 600 E3s have been identified in the human genome so far (Scheffner and Kumar, 2014). These are grouped into three main families based on structure and mechanism of action: RING/U-box E3 ligases, HECT E3 ligases and the smaller group of RING-between-RING (RBR) E3 ligases; although only the first two groups will be discussed in this introduction as RBR ligases have not yet been investigated to the same degree as RING and HECT ligases.

1.3.1. RING E3 ligases

The RING (Really Interesting New Gene)/U-box E3 ligase group (RING-type) is the largest of the three groups to exist in humans. RING E3 ligases adopt a 'cross-brace' configuration by interacting with two zinc ions via coordination bonds with ~8 cysteine/histidine residues (Figure 1.3.1A). U-box RING-types, however, assume the same conformation but without the presence of zinc, by utilising non-covalent interactions within the centre of the protein (Metzger *et al.*, 2014). The absence of intrinsic activity in RING-type E3 ligases means that they rely on the interaction and ubiquitin-binding capacity of the E2-conjugating enzyme to both facilitate ubiquitin transfer as well as determine the type of ubiquitination. While these enzymes are capable of acting as individual monomers, quite often they form active homo-/hetero-dimers. Dimerization can be mediated either via interaction of the RING domains and C-terminus or by the interaction of both N- and C-terminal regions between monomers. Homodimers tend to have two active RING-type ligases where both are capable of binding to E2~Ub. Whereas heterodimers often have one active E2~Ub binding monomer and one inactive non-E2~Ub binding monomer. The E2 binding site is created by two loops which form a groove with a central helix. This structure correctly orientates the E2~Ub conjugate and the recipient substrate to enable direct transfer of ubiquitin in one step (Berndsen and Wolberger, 2014). The first time the RING-E2~Ub intermediate was structurally solved, using an RNF4 homodimer bound to UbCH5a, revealed the structural and mechanistic basis for the E2~Ub-E3 interaction step (Plechanovova *et al.*, 2012). Binding of UbCH5a~Ub to the RNF4 homodimer caused rigidity of the E2 conformational sampling into a position where the ubiquitin C-terminus was primed for attack within an active site on the E2, and enabled interaction with the substrate lysine leading to ubiquitin transfer. This was also observed in experiments

involving the RING heterodimer BRCA1-BARD1 (Pruneda *et al.*, 2012). This activation of the E2~Ub conjugate by initiating a closed conformation ready for ubiquitin transfer to the substrate, suggests that these ligases act as more than just scaffolds and are maybe more like activators/repressors of ubiquitination. As for substrate interactions, this seems to be more complex, especially since many ligases are able to interact with the same substrate and that the recognition of substrates occurs by different means (Metzger *et al.*, 2014). Some RING-type ligases contain localisation signals to direct them towards their substrates or are embedded in the plasma membranes of organelle. Others recognise substrates based on post-translational modification such as phosphorylation or acetylation, or even require the presence of an adaptor protein. Examples include the RING E3 ligase Cbl (Figure 1.3.1B) which acts individually and binds to phosphorylated substrates via an SH2 (Src homology 2) domain. Whereas SCF complexes (Figure 1.3.1B) and the APC ligase are formed from different proteins coming together to facilitate ubiquitination (Jackson *et al.*, 2000). This diversity in RING-type E3 ligases enables the ubiquitination of a plethora of protein types and investigation of these proteins advances the understanding of ubiquitination and its role in cellular control.

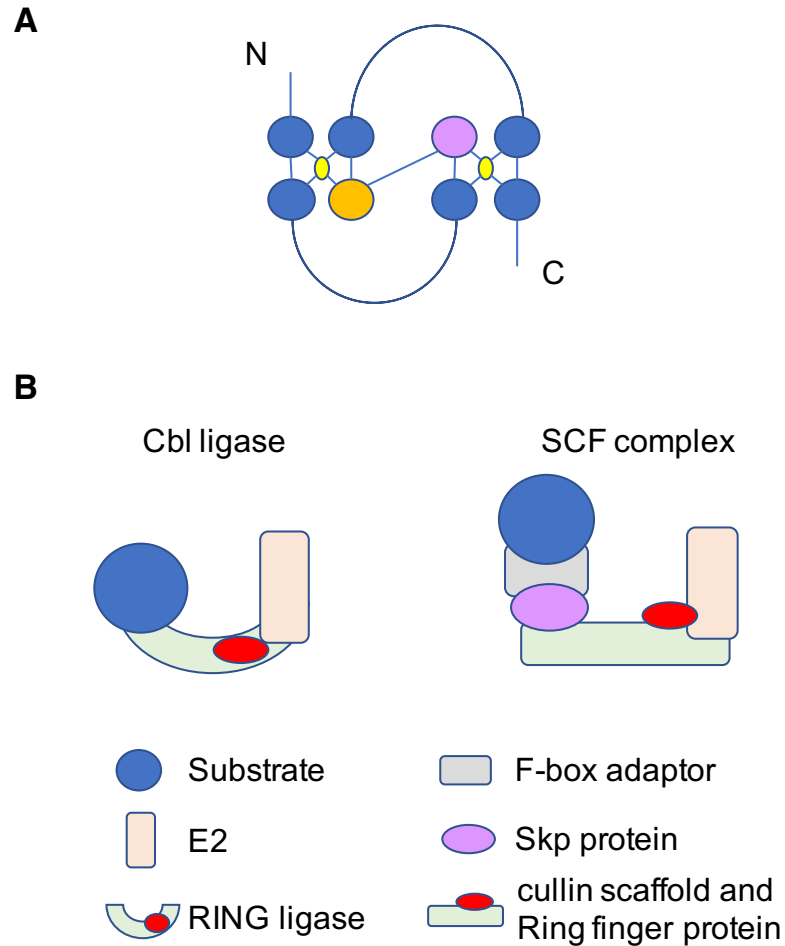


Figure 1.3.1 - Schematic diagrams of the RING domain and example ligases

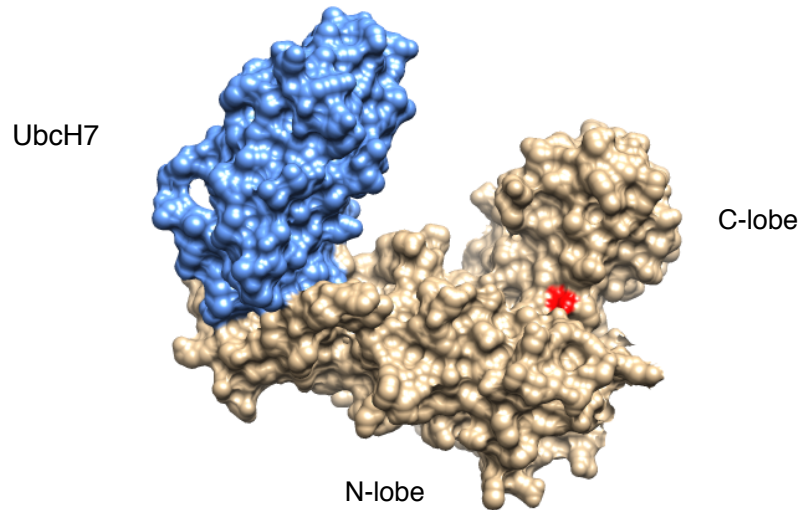
A) RING domain layout where 6 cysteine residues (blue circles), a histidine residue (purple circle) and another which can be either cysteine or histidine (orange circle) form coordination bonds to create a cross-brace conformation with zinc ions (yellow circles). N and C refer to the amino and carboxyl termini B) Representations of a single component RING ligase, Cbl (Casitas B-lineage lymphoma) (left), and the multicomponent RING ligase SCF (Skp, Cullin-1, F-box) complex (right) interacting with an E2 enzyme and substrate. A key is provided underneath the diagrams.

1.3.2. HECT E3 Ligases

In contrast to RING-type E3 ligases, HECT (Homologous to E6AP C-Terminus) E3 ligases contain intrinsic catalytic activity. There is a total of 28 HECT family E3 ligases currently known to be expressed in humans (Lorenz, 2018). These E3 ligases contain a catalytic cysteine residue within the HECT domain which creates an intermediate cysteine-ubiquitin complex making the final ubiquitin transfer a two-step reaction (Huibregtse *et al.*, 1995). The characteristic HECT domain was originally described in the protein E6AP (E6-associated protein) (Huang *et al.*, 1999). Deletion analysis identified the last 84 amino acids of E6AP as essential for p53 and E6 protein ubiquitination by E6AP. This functional region was also identified in Nedd4 establishing this region as a common trait in proteins related to ubiquitination (Huibregtse *et al.*, 1993). Structural analysis by Huang and colleagues revealed the HECT domain as bilobal in nature. Mutagenesis of the interface between the lobes interfered with catalysis (Huang *et al.*, 1999), which was similar to the mutations identified in Angelman syndrome, highlighting the region between the lobes as important to function. The catalytic cysteine residue resides in the smaller C-lobe and the E2 docking site in the larger N-lobe with a flexible hinge region connecting the two lobes. The E6AP HECT domain structure bound to UbCH7 can be seen in Figure 1.3.2A, where the HECT domain has an L-shaped conformation. The flexible hinge enables rotation of the lobes upon E2 and ubiquitin binding bringing the cysteine residues closer together (Verdecia *et al.*, 2003). The E2 binding site on the N-lobe consists of a hydrophobic region which contacts a specific phenylalanine residue in the E2 enzyme as shown by E6AP~UbCH7 (Huang *et al.*, 1999). Further analysis involving ubiquitin-charged UbCH5B interacting with Nedd4L-HECT domain construct showed that, while the contacts between UbCH5B and the N-lobe were comparable to those for the already published E6AP~UbCH7 complex, the rotation of the C-lobe closer towards the ubiquitin molecule resulted in a much more compact HECT structure as seen in Figure 1.3.2B (Kamadurai *et al.*, 2009). A distance of only ~8 Å was observed between the UbCH5B cysteine site and Nedd4L catalytic cysteine site (mutated to Ser/Ala in the experiment), although these active sites were still not close enough to facilitate ubiquitin transfer. The reorientation of the C-lobe is further supported by the many arrangements published for other HECT E3 ligases such as Smurf 2 at ~50 Å between cysteine positions (Ogunjimi *et al.*, 2005) and WWP1(WW domain containing E3 ubiquitin protein ligase 1) at 16 Å away (Verdecia *et al.*, 2003). The distance observed for Nedd4L is much smaller than those observed for other modelled HECT~E2 complexes, showing the structural conformation just before ubiquitin transfer. Collectively, these structures at

various stages of C-lobe reorientation show that rearrangement of the HECT lobes is essential for the generation of the HECT~Ub intermediate.

A



B

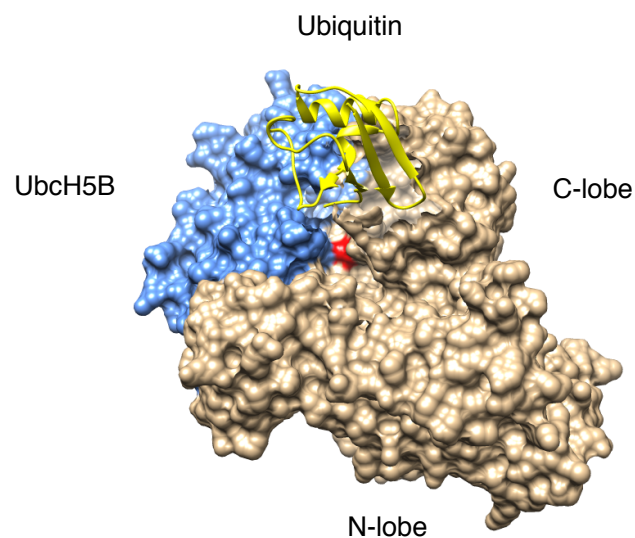


Figure 1.3.2 - HECT domain structures for E6AP and Ned4L when bound to E2/E2~Ub

A) Surface structure of the E6AP HECT domain (gold) bound to UbchH7 (blue). B) Structure of Ned4L HECT domain (gold) bound to UbchH5B (blue) and ubiquitin (yellow). The catalytic cysteine residue is highlighted in red. Structures solved by Huang et al. (1999) and Kamadurai et al. (2009) using X-ray diffraction. Images generated using Chimera.

While all HECT E3 ligases share the presence of the HECT domain for catalysis and E2 interactions, the domains in the N-terminal region from HECT are responsible for substrate specificity. There are three subgroups: the HERC family, the 'other HECT' family and the Nedd4 family (Lorenz, 2018). The HERC family has 6 members and all contain at least one RCC1 (regulator of chromatin condensation 1) domain and range from ~1000 to ~5000 amino acids in length. While HERC3-6 only contain the HECT domain and RCC1, HERC1 and 2 have additional domains in the N-terminal regions linked to substrate interactions. In contrast, the 'other HECT' family only share similarity via the HECT domain and can contain varying lengths and domains outside of HECT. Lastly, the Nedd4 family of HECT E3 ligases, of which there are nine family members, all contain very similar domain architecture but with varying sequences and linker lengths. This Nedd4 sub-family of ubiquitin ligases is the centre of this work and will be discussed further in the next section.

Errors in ubiquitination as a result of aberrant regulation of E3 ligases have been identified in many diseases, such as neurological, inflammatory and cardiac disease. For example, mutations in the gene encoding the RBR ligase parkin have been linked to the development of Parkinson's disease (Charan and LaVoie, 2015). Loss of function, a decrease in solubility and aggregation of parkin results in reduced ubiquitination of downstream substrates such as mitochondrial surface proteins. The lack of parkin enzymatic activity on the mitochondrial surface prevents recognition and clearance of mitochondria via the autophagosomal degradation pathway. In turn, this results in the toxic build-up of dysfunctional mitochondria and neurone death (Karbowski and Youle, 2011). The role of parkin in neurodegeneration is further supported by parkin's ability to protect neurones during times of apoptotic stress, ER stress and dopamine stress. Since the structural solution of parkin (Riley et al., 2013; Wauer and Komander, 2013), further understanding of the molecular mechanisms and functions of parkin has pushed forward the idea of this E3 ligase being a potential target for agonistic or allosteric activator design in disease treatment. Additionally, parkin has also been shown to play a role in cardiac injury after myocardial infarction (Willis et al., 2015). The recruitment of parkin by the ubiquitin ligase Mfn2 (mitofusin-2) in cardiomyocytes is required for the removal of damaged mitochondria in order to control cellular homeostasis during reperfusion and tissue recovery. Parkin also displays a protective element in cardiac tissue to help minimise damage should a second event occur. Thus, reduced levels of functional parkin can result in heart disease and failure to develop a protective response upon a following heart attack.

Another neurological disease called Angelman syndrome has been linked to deletion or point mutation of the gene *UBE3A*. *UBE3A* encodes the HECT E3 ligase E6AP with substrates identified in many different regulatory pathways, such as the kinases CDK1 and CDK4 involved in cell cycle regulation, and MAPK1 in cell proliferation (Cooper et al., 2004). The abnormal regulation of these proteins, along with ~120 other substrates, due to loss of E6AP enzymatic activity, results in symptoms such as seizures, loss of speech and mental deficiencies. In contrast to E6AP, the loss of another HECT E3 ligase called Itch, leads to chronic lung inflammation, enlarged liver and spleen as well as autoimmune disease (Venuprasad et al., 2015). The loss of Itch results in stabilisation of the protein JunB, leading to an imbalance in T helper cell differentiation towards Th2 type cells. The increase in Th2 cells results in significantly increased allergic responses in Itch deficient mice, further supporting the role of Itch in the regulation of inflammatory response. Collectively, these examples show the importance of having functional E3 ligases and the implications mutation and/or loss of regulated ubiquitin ligase activity can have in the development of disease.

In addition to the above diseases, E3 ubiquitin ligases have also been identified as drivers of oncogenesis, by either being a proto-oncogene themselves or by overexpression leading to a decrease in tumour suppressor proteins. Further understanding of the structure-activity relationship of E3 ligases and their substrates has led to the development of novel cancer therapies. For example, RING E3 ligase overexpression has previously been associated with cancer growth. One of the most studied RING E3 ligases is Mdm2. Mdm2 overexpression has been found in many cancer types, including prostate cancer (Leite *et al.*, 2001). This E3 ligase interacts with and regulates the activity of the tumour suppressor protein p53 via three methods – ubiquitination of p53 leading to degradation; binding and blocking of the p53 activation domain; and induction of nuclear export preventing DNA interaction of p53. These all result in a reduction of apoptosis and cancer cell survival (Zhao *et al.*, 2015). Successful development of inhibitors of Mdm2 has shown increasing capabilities in disruption of the protein-protein interaction restoring tumour suppressive protein activity, in turn reducing tumorigenesis. This wouldn't have been possible without the knowledge of the structure, interaction and function of these enzymes and substrates. This is not the only the case as there is an accumulating body of evidence supporting the association of HECT E3 ligase overexpression as well in the development and progression of cancer (Bernassola *et al.*, 2008). The focus of the work in this thesis will be on the

Nedd4 superfamily of HECT E3 ligases with most interest focused on the E3 ligase WWP2.

1.4. Nedd4 family of HECT E3 ligases

The human genome is known to encode 9 HECT E3 ligases that make up the Nedd4 (neural precursor cells-expressed developmentally down-regulated 4) superfamily, including HECW1, HECW2, Nedd4, Nedd4-2 (Nedd4L), Smurf 1, Smurf 2, Itch, WWP1 and WWP2. The Nedd4 family members differ in size from 748 residues up to 1606 but have a well-characterised structure consisting of three main regions. Situated at the N-terminus is the C2 domain, in the middle resides up to four double tryptophan (WW) domains, and at the C-terminal end sits the HECT domain (Scheffner and Kumar, 2014; Soond and Chantry, 2011; Yang and Kumar, 2010). The domain organisation can be seen in Figure 1.4.1 outlining all nine members of the Nedd4 family.

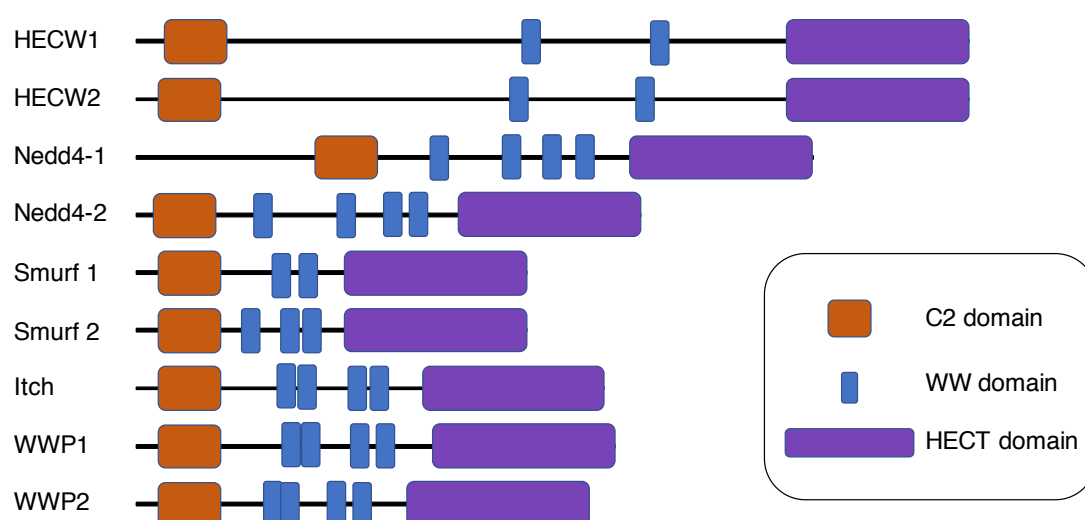


Figure 1.4.1 - Members and domain arrangements of the Nedd4 E3 ligase family

Schematic representations of the domain architecture of Nedd4 family E3 ligases as described by uniprot. N-terminal C2 domain is shown by the orange box, WW domains are represented as blue boxes and the purple box at the C-terminus is the catalytic HECT domain.

The C2 domain folds into an eight stranded β -sheet sandwich and interacts with intracellular calcium via conserved aspartic acid side chains in the β 2/3 and β 6/7 loop regions of the sandwich (Sutton *et al.*, 1995). This domain facilitates the localisation and interaction with membrane phospholipids (Knopf *et al.*, 1986), and regulates many

proteins using the fluctuations in calcium as a mediator of protein translocation. Plant and co-workers (1997) showed that removal of the C2 domain from Nedd4 abolished its localisation to the plasma membrane and phospholipid interaction. The presence of the C2 domain in Nedd4 is, therefore, required for directing the activity of this ligase at cellular membranes and may help the targeting of these ligases towards particular substrates as implied by its conservation in all 9 members. More recently, solution structure of the Smurf 2 C2 domain highlighted that this domain adopts the same folding pattern common to C2 domains. However, there was an upside-down orientation with the N- and C-termini at the 'bottom' of the structure (class II), unlike the C2 domain of PKC with these termini at the 'top' (class I) (Wiesner *et al.*, 2007). Based on sequence similarity across the Nedd4 family, this suggested a conserved class II C2 domain in this group. Further investigations into the Smurf 2 C2 domain to understand more about its role in Smurf 2 activity uncovered interactions with the HECT domain. The presence of the C2 domain led to steady-state levels of Smurf 2 with no ubiquitination activity however removal of the C2 domain resulted in auto-ubiquitination activity of the ligase. This supports the function of the C2 domain in regulating Smurf 2 activity and provides a possible mechanism employed by other Nedd4 E3 ligases. On this note, the same role of C2 has been reported for Nedd4 (Wang *et al.*, 2010), highlighting this region as important in the control of ubiquitination and auto-ubiquitination for these members.

Alternatively, both WW domains and linker regions have shown auto-inhibitory functions in Nedd4 family proteins. Itch showed autoinhibition could be relieved by the addition of NDFIPs (Nedd4 family interacting proteins) which bind to WW domains (Riling *et al.*, 2015). Further study revealed intramolecular interactions of the WW2-3 region with the HECT domain as the mechanism of autoinhibition in which mutation of the 2nd tryptophan, important for PPxY recognition, resulted in Itch activity similar to NDFIP interaction. While this was also thought to apply to the closely related WWP2 enzyme, evidence by Chen and co-workers (2017) showed that it is the linker between WW2 and WW3 which is involved in WWP2 inhibition. The linker forms a helix that interacts with both the N- and C-lobes of the HECT domain with the last 6 residues of WW2 also contacting the HECT domain (Chen *et al.*, 2017). This region of WW2 and the N-terminal residues of the 2-3 linker are proposed to interact with a ubiquitin 'exosite' in the HECT domain, previously identified in Nedd4 and Smurf 2 (Maspero *et al.*, 2011; Zhang *et al.*, 2016), but yet to be formally located in WWP2. These mechanisms employed by Nedd4 E3 ligases result in tight regulation of their activity as

well as providing insight into possible ways to limit the activity of these proteins when they are over-active.

WW domains within the central region of the proteins vary in number within the Nedd4 family of proteins with only 2 residing in HECW1, HECW2 and Smurf 1, three within Smurf 2 and the rest containing four WW domains. WW domains are usually 30-40 amino acids in length with two distinctive tryptophan residues towards either end of the domain interspaced by ~20 residues. The second tryptophan residue, however, is sometimes replaced by tyrosine/phenylalanine, as seen for example in WW4 of WWP2 and Smurf 1 WW2 domain. The Nedd4 family of ligases are typical class I WW domains with a binding preference for the proline-rich PPxY residue sequence. There are, however, a further 3 classes of WW domain. These other classes can interact with PPLP motifs (class II), proline-arginine (class III) and phospho-serine-proline (class IV) (Bedford *et al.*, 2000; Espanel and Sudol, 1999). WW domains both within the same protein and between different proteins can share the same interaction partner, yet also show a selective preference towards particular targets. As the motif is the same, this shows that the sequence outside of the proline motif is key for determining the binding profile of the individual WW domains.

Several WW domains within Nedd4 E3 ligases have been structurally solved. This has helped to uncover atomic aspects of WW domain folding and the interactions with target substrates (Aragon *et al.*, 2011; Aragon *et al.*, 2012; Chen *et al.*, 2017; Chong *et al.*, 2006; Chong *et al.*, 2010; Kanelis *et al.*, 2001; Panwalker *et al.*, 2016; Qi *et al.*, 2014). WW domains all share a 3 stranded anti-parallel β -sheet conformation with disordered N and C terminals situated either under or heading away from the secondary structures. Curvature of the strands creates a central groove where substrates can interact. Within this twisted region resides a group of conserved residues that form a hydrophobic patch. Two key residues are a tyrosine/phenylalanine on the second strand and the C-terminal tryptophan residue that make up the XP pocket. The XP pocket specifically orientates the first two prolines of the PPxY motif in the substrate, promoting interaction between the two proteins (Marcias *et al.*, 1996). A second specificity pocket is formed by another two residues - a hydrophobic amino acid often leucine, valine or isoleucine located in the second strand and a histidine residue in loop II. These are involved in contacting the tyrosine residue of the PPxY motif. Mutagenesis highlighted the hydrophobic residue as the major contributor towards interactions of class I WW domains, as mutation of this residue abolished tyrosine binding and changed the WW

domain recognition towards the class II PPLP substrate type. Histidine showed more diversity in which 5 substitutions were possible that retained PPxY binding (Espansel and Sudol, 1999). As the XP pocket, 2nd specificity pocket and PPxY motif are shared across the Nedd4 family, the specific orientations, residue type and other contacts must relay the specificity shown by WW domains towards substrates. Uncovering this information in a variety of WW domains with similar and unique interaction partners should reveal information about the specific interface between Nedd4 E3 ligases and their specific target proteins.

Comparison of the XP residues in the bound structures for Smurf 1 WW2 (PDB: 2LTX) Smurf 2 WW3 (PDB:2LTZ) and Nedd4-2 WW2 (PDB: 2LTY) (Aragon *et al.*, 2012) in the presence of the same substrate Smad 7 peptide is shown in Figure 1.4.2. While the first XP residue is a tyrosine in all three domains, only Nedd4-2 contains a canonical 2nd tryptophan residue, whereas both Smurf 1 WW2 and Smurf 2 WW3 have a phenylalanine residue at this position. Nevertheless, similar orientations of the XP residue side chains are observed when in contact with the first and second proline residues of Smad 7. Variation is present, however, in the structural interfaces between these domains. A tight hairpin fold is observed when the peptide interacted with Nedd4-2, whereas for Smurf 1 and Smurf 2 WW2 and WW3 domains, respectively, this turn is not nearly as tight. Other residues in Nedd4-2 WW2 that contact the Smad 7 peptide are situated across the whole domain. R389 and T391 on the third strand, along with V384 and H386 on the second strand interact with the conserved tyrosine residue in the PPPY motif, and the additional residue T391 in the WW domain also contacts the PPPY prolines. Further away from the proline motif in the Smad 7 peptide, additional contacts are made with L378 and R380 in loop I-strand 2 of the WW domain. Mutation of these residues in Nedd4L WW2 resulted in a decrease in affinity towards the Smad 7 peptide. In contrast, no loop I contacts are present for Smurf domain interactions, but mutagenesis of the arginine residue at the equivalent Nedd4-2 WW2 R380 position in Smurf 1 WW2 (R295) and Smurf 2 WW3 (R312) had the same decrease in affinity, showing the role of this residue in Smad 7 peptide interaction. The conformation of the substrate and the contacts differ between domains highlighting the complexity in understanding the preferences and interactions between WW domains and substrates. Analysis of Nedd4-2 domains interacting with Smad 2 and Smad 3 showed that interactions were dependent on the activation of the Smad protein via phosphorylation of a threonine residue upstream of the PPxY motif and involved more than one WW domain (Gao *et al.*, 2009). This brings in the idea that both modifications

and tandem domain functionality can also control WW domain interactions with target substrates, and that particular substrate sequences outside of the PPxY motif may determine individual WW domain: substrate preferences, as shown by the constitutive Smad 7 binding to Nedd4-2 in contrast to Smad 2/3 interactions.

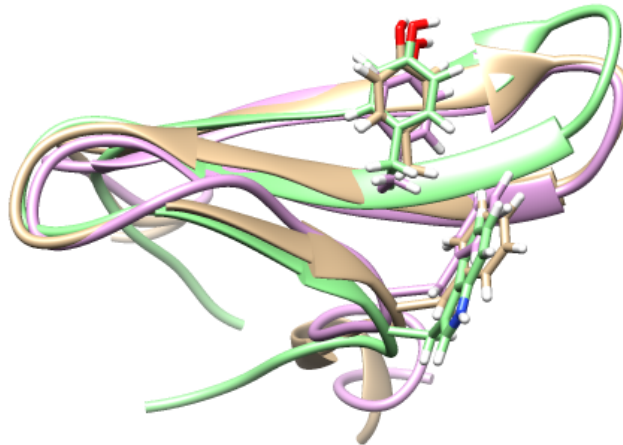


Figure 1.4.2 - Comparison of the XP residues of Smad 7 bound WW domains in Nedd4 family members

The structures of Smurf 1 WW2 (gold) (PDB: 2LTX), Smurf 2 WW3 (purple) (PDB: 2LTZ) and Nedd4-2 WW2 (green) (PDB: 2LTY) when bound by Smad 7 peptide (Aragon *et al.*, 2012). The equivalent first XP pocket tyrosine residues on the second strand and the tryptophan/phenylalanine that constitutes the second residue of the XP pocket are shown for all three domains. Comparison and image generation carried out using Chimera.

Another example where WW domains show selective preferences for substrates can be seen in the founding family member, Nedd4. Staub and colleagues (1996) identified Nedd4 as an ENaC (epithelial sodium channel) binding partner via two PY motifs present in the C-tail of the membrane channel leading to internalisation of the channels and ubiquitin-dependent degradation. Other Nedd4 family members such as Nedd4-2 and WWP2 also target ENaC (McDonald *et al.*, 2002). The regulation of ENaC by these ligases is important as seen in Liddle's syndrome, where loss or mutation of the PY motif leads to reduced ubiquitin-dependent degradation and the subsequent increase in ENaC protein levels leading to the development of hypertension (Goulet *et al.*, 1998; Harvey *et al.*, 1999). GST pull-downs using WW domains of WWP2 and Nedd4-2, and ENaC subunits showed all WW domains except WW1 of WWP2 could interact with ENaC. Similar to WWP2, all domains of human Nedd4 bound to ENaC, but not the first

WW domain compared to rat and mouse homologues (Lott *et al.*, 2002). Only a single amino change from an arginine to a threonine on the third strand was capable of enabling human Nedd4 WW1 to interact with ENaC. Analysis of the interface between the Nedd4 WW domains and ENaC found that WW3 had the highest affinity for ENaC, and the XP pocket residues phenylalanine 476 and tryptophan 487 displayed the same accommodation of the two prolines of the PPxY motif (Kanelis *et al.*, 2001) as described for other WW domains. Unlike the Smad 7 motif which has another proline at the x amino acid site, the x residue of the ENaC motif is an asparagine but forms the same polyproline helical conformation. The contacts with the tyrosine are made by an isoleucine and histidine residue in WW3, with additional contacts to a leucine residue C-terminal of the tyrosine in ENaC (Kanelis *et al.*, 2001). Investigations into the specific contacts in these interactions is showing that there are individual residues that are outside of the classic motifs that are dictating specificity and only by exploring these individual contacts can we begin to understand how these ligases function.

mRNA screening showed wide distribution of HECT E3 ligases across cell types, yet certain ligases showed selective over expression in cancer cell lines (Komuro *et al.*, 2004). For example, Nedd4, WWP1 (Komuro *et al.*, 2004) and WWP2 (Soond *et al.*, 2013) have shown over expression in melanoma, prostate and breast cancer. Komuro also showed higher levels of Smurf 2 in colon, breast and ovarian cancer, whereas WWP1 was over expressed in colon, stomach and an ovarian cancer cell line. Expression analysis by Kuratomi *et al.* (2005) also identified Nedd4-2 over expression in several cancer cell lines including colon, breast, ovarian and melanoma. More recently, analysis of WWP2 isoforms in patient samples of these cancers revealed potential stage specific and cancer type specific expression levels (Soond *et al.*, 2013). This analysis, however, has only involved a small sample size thus further study is required to determine if this can be observed across more cases. This would be interesting to find out as it may provide another prognostic marker for these cancers, because, for example in prostate cancer, there are cases where the biomarker PSA is not sufficient in identifying cancer (PSA negative prostate cancer cases) unless other symptoms are present (Birtle *et al.*, 2003).

1.4.1. WWP2 and its isoforms

WWP2 was first discovered in 1997 (Pirozzi *et al.*, 1997; Harvey and Kumar, 1999) and was identified as a member of the Nedd4 family. The full-length protein consists of 870 amino acids which form independent domains as shown in Figure 1.4.3. WWP2 has

not been studied as much as other Nedd4 family members, however research is starting to answer some of the unknown aspects of this protein. WWP2 has 3 main isoforms found in humans generated from alternative splicing and/or use of alternate promoters. WWP2-N is created from the retention of an intron between exon 9 and 10 that contains an alternative stop codon resulting in a premature stop signal and translation truncation. This results in a shorter protein consisting of only amino acids 1 – 335 of the N-terminal region and contains the C2 domain and the first WW domain. In contrast, another isoform WWP2-C consists of the C-terminal region from residue 440 – 870 containing WW4 and the HECT domain. This is created from an internal putative promoter at the start of exon 13. Other isoforms have been found in human cells, however further identification and investigations are required to determine if these are true isoforms and have a functional role.

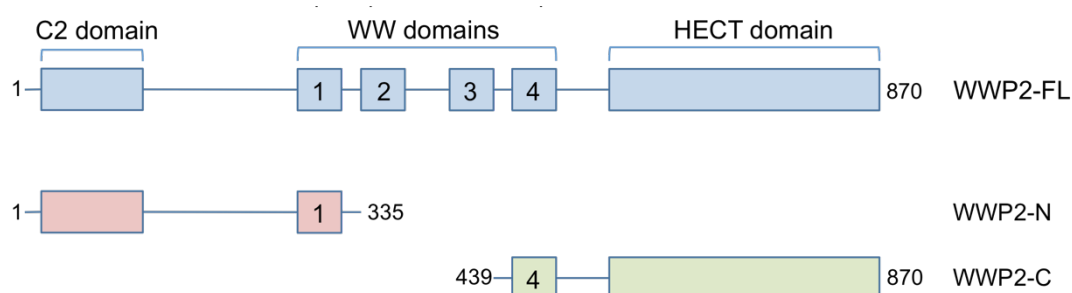


Figure 1.4.3 - Diagrams defining the regions present in WWP2 isoforms

A summary of the three well-known active isoforms of WWP2, and their individual domain architecture – N-terminal C2 domain important for calcium and phospholipid binding, central WW domains involved in substrate interactions and the characteristic HECT domain at the C-terminus. WWP2-N contains only the C2 domains and WW1 regions, whereas WWP2-C consists of WW4 and the HECT domain. Numbers are the start and end refer to the amino acid number in relation to full-length WWP2.

Structural solution of the domains within WWP2 is currently limited to the HECT domain, WW2 and the WW2-3 linker region. Crystallisation of the HECT domain (Gong *et al.*, 2015) showed the same bilobal T-shape conformation as other HECT domains such as WWP1 (Verdecia *et al.*, 2003), Itch and Nedd4-2, which contrasts with the L-shape of Nedd4 (Maspero *et al.*, 2011) and Smurf 2 (Ogunjima *et al.*, 2015). These different conformations are due to the flexibility of the hinge region that connects the

two lobes, and most likely the differences in intra-molecular contacts of the two lobes. Nedd4 was suggested to convert to the more stable T-shape conformation when interacting with the E2~Ub complex (Maspero *et al.*, 2013), however, evidence of the yeast HECT E3 ligase Rsp5 suggests that the L-shape conformation can be maintained even when bound by Ub and substrate Sna3 (Kim *et al.*, 2011). E6AP also showed an L-shape conformation (Huang *et al.*, 1999). Perhaps these conformations are one way in which the individual HECT domains separate their activity from other members of the same family, and it may be useful to untangle the specifics for ways to target selective HECT domains. The T-shape conformation of the WWP2 HECT domain was further confirmed by Chen *et al.* (2017) as its autoinhibited form by interaction of HECT with the WW2-2-3 linker region. Chen and co-workers also showed that WW2 of WWP2 reflected the 3 stranded anti-parallel β -sheet previously identified for WW domains. Unbound structures for the other WW domains within WWP2 as well as substrate-bound structures however, have yet to be solved.

1.4.2. WWP2, signaling pathways and their role in prostate cancer

The most common cancer to affect men is prostate cancer (PCa), with around 47,000 men being diagnosed each year in the UK (PCUK, 2017). The prostate is a gland that sits beneath the bladder. There are three main regions of the prostate – the central zone which surrounds the ejaculatory duct which connects the seminal vesicle to the urethra; the transitional zone that is located directly below the bladder and surrounds the urethra; and the peripheral zone which is at the posterior region of the gland (Hammerich *et al.*, 2008; Packer and Maitland, 2016). The acinus of the prostate is lined with secretory luminal cells that are important for the release of seminal fluid into the duct. The luminal layer is surrounded by further layers of progenitor basal cells and stem cells additionally encased by a stromal cell environment (Packer and Maitland, 2016). The stem cells divide into either more stem cells or cells that differentiate into basal cells in order to maintain the basal cell and luminal cell layers. To date, the heterogeneity of the prostate gland has restricted the identification of specific factors associated with PCa development, along with no obvious epidemiological factor.

PCa is grouped into three main stages based on whether it has spread from the prostate or how far it has spread (NICE clinical guidelines, 2014). Localised PCa resides only in the prostate gland itself, whereas localised advanced has spread outside of the prostate to the nearby tissues such as the bladder and seminal vesicle. The last stage is metastatic PCa in which the cancerous cells have spread further into

the body, often through the lymphatic system to the bones. In addition to this, a further refined grading system called the Gleason score is used to describe the cellular histology of the prostate tissue rather than just location for the assessment of PCa, ranging from 1 to 5 (Hammerich *et al.*, 2008; Packer and Maitland, 2016; PCUK, 2017). Due to the heterogeneity of PCa, two values are given in order to represent the primary and secondary tissue grades present in patient samples. These resulting scores can therefore range from 2 (1 + 1) to 10 (5 + 5), where the individual grades of 1 and 2 refer to non-cancerous tissue while 3, 4 and 5 indicate a cancerous profile. Grades of 6-10 (3+3 to 5+5) rank the cancer from slow growing and unlikely to differentiate further into aggressive disease, up to advanced metastatic disease that is aggressive in nature and spread outside of the localised region of the prostate. PCa can form several foci in the gland, meaning that increased samples are taken by biopsy of the prostate in order to determine the presence or absence of PCa.

There are currently many treatment options based on the Gleason score and stage of the cancer. Both watchful waiting and active surveillance are used either when men do not want to undergo treatment or when PCa is localised and has a low or intermediate risk of future progression. The most well-known active treatment is androgen ablation therapy. These therapies act upon the androgen receptor signaling pathway by either chemical or surgical means in order to reduce serum hormone levels to that of castration conditions (Gomella *et al.*, 2014). Other treatment options include radiotherapy and chemotherapy, however, each of these treatments can come with side effects which result in a poor quality of life. While there are many options, the major problem associated with hormone therapy and other current treatments in the long term is that, in most cases, the tumours return in ~18 months as hormone-independent PCa (also known as castrate-resistant PCa) which is more aggressive and quick growing (Gomella *et al.*, 2014; Singh and Kim, 2016). Hence, therapies which target other molecular factors associated with AR-independent cancer for use in castrate-resistant PCa are required.

Nan and colleagues (2003) showed PTEN as a negative regulator of the androgen receptor, a nuclear steroid receptor associated with PCa development, via AKT signaling suggesting PTEN as a target for PCa-associated therapeutics. The prevalence of PTEN loss in castrate-resistant cancer however is high suggesting a different drug target involved in PTEN regulation may be a better strategy for treating advanced PCa (Carver *et al.*, 2011; Taylor *et al.*, 2010). PTEN loss has been linked to

over expression of HECT E3 ligases in PCa, as well as increased loss of other tumour suppressor proteins such as Smad proteins in the canonical TGF- β signaling pathway. Hence the ubiquitination pathway presents an ideal target for further investigation, especially as members of the Nedd4 family of E3 ligases have substrate specific WW domains attractive for drug design, which will be explored further in this work.

Over expression of WWP2 isoforms has been linked to the development and progression of prostate cancer. Expression profiling in patient samples found a stage specific expression pattern in prostate cancer (Soond *et al.*, 2013). While this study used a small sample size, further investigation may reveal WWP2 as a potential biomarker for prostate cancer, particularly important in PSA negative and PTEN negative cases. Furthermore, the effect of WWP2 overexpression on tumour suppressors such as Smad proteins and PTEN has significant effects on cancer cell abilities to bypass usual control mechanisms such as growth inhibition and apoptosis (Chantry, 2011). This highlights the importance of regulation of this ligase in the suppression of tumour formation.

Smad proteins in the TGF- β signalling pathway (more detail provided in section 1.6) and the lipid phosphatase PTEN (further described in section 1.7), are known binding partners of WWP2 isoforms and have been linked to cancer development (Chen and Matesic, 2007; Soond *et al.*, 2013). WWP2 isoforms have been shown to respond differently in the presence of TGF- β . Levels of WWP2-FL increased with TGF- β stimulation with the highest level detected on day 2 which then became undetectable on day 4, whereas WWP2-N was only detectable at low levels for the first 3 days and WWP2-C undetectable throughout (Soond and Chantry, 2011). This suggests that the regulation of WWP2 isoform expression may be regulated in a context-specific manner, but the specifics have yet to be fully elucidated. Analysis of WWP2 isoform interactions showed that these differences may be mediated by the interaction and ubiquitination of Smad proteins that act as the secondary messengers in TGF- β signaling. Differential binding of WWP2-FL to both Smads 2 and 3 compared to Smad 7 was observed, although TGF- β stimulation increased interaction with Smad 7 resulting in greater Smad 7 degradation. On the other hand, WWP2-N bound preferentially to Smad 3 under both TGF- β negative and positive conditions, with reduced binding to Smad 2 only observed with TGF- β stimulation. Interaction with Smad 2 and 3 in TGF- β positive conditions is thought to stabilize the cellular concentrations of these transcription factors as WWP2-N lacks the catalytic HECT domain. The binding of Smad 2 by

WWP2-FL was also only observed under TGF- β stimulation. WWP2-C bound to Smad 7 in TGF- β -dependent conditions to a greater degree than WWP2-FL. Due to the isoform architecture, it was inferred that WW1 targets R-smads 2 and 3, and WW4 selectively binds inhibitory Smad 7. More recent NMR analysis highlighted WW4 as capable of binding to Smads 2 and 3 as well, although the highest affinity was towards phospho-Smad 7 followed by Smad 3 (Wahl, 2016).

Soond and Chantry (2011) also found that the isoforms were capable of modulating WWP2-FL activity. While auto-ubiquitination activity of WWP2-FL was observed in the presence of WWP2-N and WWP2-C, only WWP2-N was detected in pull-down analysis. Binding of WWP2-N to WWP2-FL under TGF- β negative conditions resulted in a decrease in Smad 2/3 levels but no change in Smad 7. However, upon ligand stimulation, the WWP2-FL: WWP2-N isoform complex dissociated and WWP2-FL preferentially targeted Smad 7. WWP2-N therefore modulates WWP2-FL autoinhibition and isoform activity in a TGF- β mediated manner.

Another interaction partner of WWP2 as mentioned above is PTEN (Maddika *et al.*, 2011), discussed later in section 1.7. The binding region in PTEN was isolated to within the phosphatase domain, suggesting WWP2 binding acts to physically prevent PTEN activity. Furthermore, the use of a catalytically inactive WWP2 abolished PTEN ubiquitination supporting WWP2 as a PTEN specific interaction, also supported by WWP2 knockdown affecting PTEN abundance. The binding site within WWP2 and its isoforms, however, have yet to be isolated as PTEN lacks the characteristic PPxY motif found in known substrates of WWP2. In order to specifically disrupt selective interactions between WWP2 and its substrates, further understanding of the protein-protein contacts involved is required.

1.5. TGF- β signalling

TGF- β (transforming growth factor- β) signaling has many roles in both the embryo and in adult systems. Smad proteins are the secondary messengers in the TGF- β family of signaling pathways that relay the signal from the cellular membrane receptors to the DNA in the nucleus to control transcription. Cytokines including TGF- β , BMPs (bone morphogenetic proteins) and activins bind to kinase receptor complexes leading to downstream signaling events. Interactions of the type I and type II TGF- β receptors is triggered by the activation of cytokines in the extra-cellular matrix which binds to type II receptors which recruits type I receptors to oligomerize with the type II (Croxford *et al.*, 2017; David and Massague, 2018). This results in phosphorylation of the type I receptor cytoplasmic GS (glycine-serine rich) domains. This activates the kinase activity of the type I receptors and induces phosphorylation of receptor-regulated smad proteins (R-smads) such as Smad 2 and 3 in canonical TGF- β signaling. Activation of these messengers by phosphorylation enables heterotrimeric complex formation with the common-mediator smad (co-smad), smad 4. This complex can then translocate into the nucleus and regulate transcriptional activity via interaction with co-activators (e.g. p300 and CBP), co-repressors (e.g. SnoN) (Soond and Chantry, 2011), or other transcription factors into a high-affinity DNA binding complex. A negative regulator of this pathway is called an inhibitory smad (I-Smads) protein, an example being Smad 7 in canonical TGF- β signaling. I-Smads inhibit downstream activity of the TGF- β pathway via competitive binding to the type I receptors (Hayashi *et al.*, 1997), by interaction with R-smads and the recruitment of specific E3 ligases to R-smads and cell surface receptors (Kavsak *et al.*, 2000; Kuratomi *et al.*, 2005; Yan *et al.*, 2016). A simplified summary of canonical TGF- β signaling is shown in Figure 1.6.1.

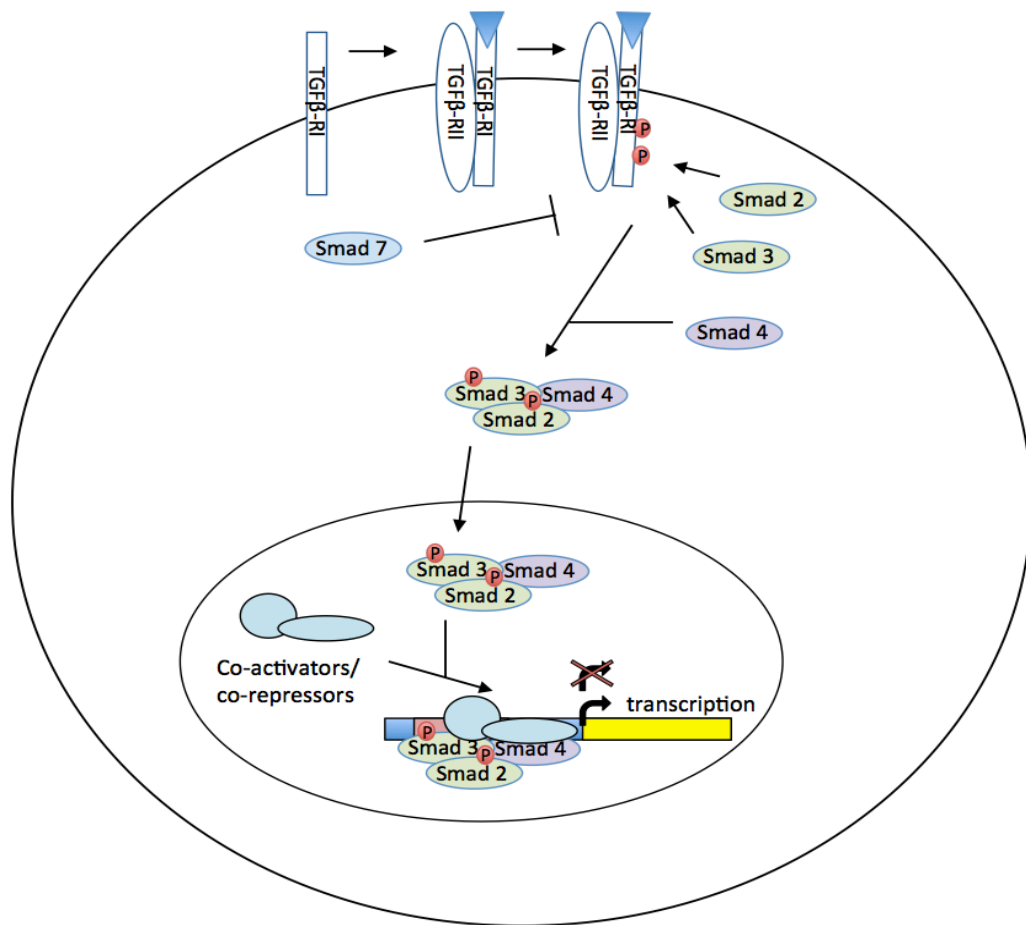


Figure 1.6.1 - Schematic representation of canonical TGF- β signaling

Activation and binding of TGF- β ligand (blue triangle) to TGF- β -RI (white rectangle) and TGF- β -RII (white circle) receptor pairs leads to phosphorylation (red 'P' circles) of the type I receptor and activation of the receptor kinase activity. TGF- β -RI recruits and phosphorylates R-smads 2 and 3 (green circles), which then oligomerise with co-Smad4 (purple circle) into a trimeric complex. The trimer then localises to the nucleus, binds to other transcription factors (pale blue circles) before DNA binding to either activate or repress transcription. I-smad 7 (blue circle) negatively regulates the TGF- β receptors and R-Smad interactions and phosphorylation. TGF- β – transforming growth factor beta; RI/II – receptor I/II; P – phosphate.

Smad proteins have different domain compositions related to their role in the signaling pathway and can be split into three groups. A common domain across all three types of smad protein is the MH2 (Mad homology 2) domain. This domain is required for mediating protein-protein interactions with other transcription factors and receptors. At the very end of this domain in R-smads, the C-terminus contains a Ser-x-Ser motif that is recognized and phosphorylated by TGF- β -RI, in turn activating the protein and

facilitating MH2 domain interactions with other Smad proteins. Next to this MH2 domain is a flexible linker region. The linker region contains sequences that can both up-regulate and down-regulate activity. Phosphorylation sites such as those for CDK8/9 (cyclin-dependent kinases 8/9) increase the function of these smad proteins in mediating transcription (Alarcón *et al.*, 2009), whereas when GSK3 (glycogen synthase kinase 3) phosphorylates the linker region, this recruits HECT E3 ligases to the transcription factor to mark it for degradation (Aragón *et al.*, 2011; Gao *et al.*, 2009). HECT E3 ligases are able to interact with these transcription factors via recognition of another motif in the linker region, a proline rich PPxY motif. This motif is conserved in all smad proteins indicating an important sequence for the regulation of these proteins. R-Smads and co-smads also contain another globular domain at the N-terminus, an MH1 domain. This domain functions as a DNA binding domain that recognizes the Smad binding element 5'-CAGAC-3' and is highly conserved in these two types of Smad protein (Dennler *et al.*, 1998). In contrast, inhibitory smads lack the MH1 domain and instead have an extended linker region that is more diverse but still retains the PPxY motif.

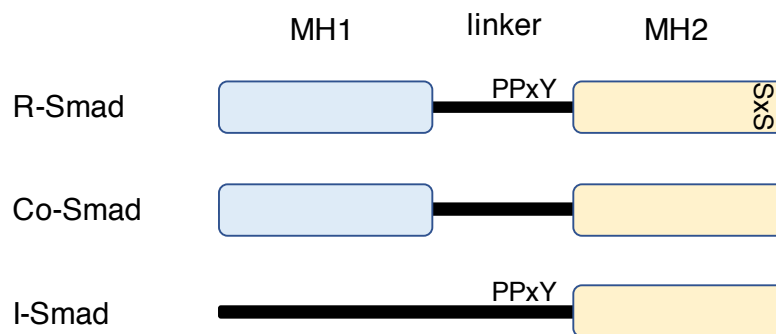


Figure 1.6.2 - Schematic representations of Smad protein domains

A summary of the characterizing domains of the different types of Smad protein. The N-terminal MH1 domain (blue) is connected to an MH2 domain (yellow) in R-Smads and Co-Smad by a linker region (black). In contrast I-smads contain the MH2 domain but lack the MH1. The PPxY motif is shown in linker region of R-Smads and I-Smads that enables E3 ligase recognition. MH1 and MH2 – Mad homology domains 1 and 2; R-smad – receptor-regulated smad; Co-smad – common-mediator smad; I-smad – inhibitory smad; PPxY – proline-proline-(any amino acid)-tyrosine; SxS – serine-(any amino acid)-serine.

Regulation of TGF- β signaling is essential as this pathway controls the activity of many genes associated with cell growth and differentiation. One mechanism employed includes self-regulation via intra-molecular contacts between the MH1 and MH2

domains of R-smad 2 and co-smad 4 (Hata *et al.*, 1997). This prevents early oligomerization and activation of these transcription factors when the pathway receptors are inactive. Phosphorylation of the C-terminal tail relieves the autoinhibited conformation between the N and C terminal regions and activates these smad proteins for downstream signaling.

As briefly mentioned earlier, Smad 7 acts as a negative regulator of TGF- β signaling. Smad 7 can down-regulate activity by many different mechanisms. Hayashi *et al.* (1997) showed that Smad 7 transiently interacts with the type I TGF- β receptors via the MH2 domain at the C-terminus of Smad 7, thereby blocking R-smad access to undergo phosphorylation by the receptor which antagonizes downstream events by R-Smads. Additional activity as a consequence of Smad 7 association with TGF- β -RI includes the recruitment of both phosphatases and E3 ligases (Ebisawa *et al.*, 2001; Kavsak *et al.*, 2000; Komuro *et al.*, 2004; Kuratomi *et al.*, 2005). This often results in silencing of the kinase domain activity of the receptors and/or internalization and degradation. More recently, evidence of direct interaction between I-Smad 7 and R-smads via the Smad 7 MH2 domain shows competitive inhibition by preventing co-Smad 4 binding. In turn, recruitment of E3 ligases such as Nedd4-2 to this complex results in ubiquitination and degradation of the R-smad proteins (Yan *et al.*, 2016). Mutation of the PPxY motif in Smad 7 did not prevent binding between Smad 7 and R-smads but did affect the recruitment of E3 ligases for R-smad turnover. Therefore, not only does Smad 7 act on the membrane receptors to inhibit TGF- β signaling but also on the R-smads themselves in turn leading to the turnover of both the receptor and activated messengers.

While Smad 7 is one regulator of the pathway via the recruitment of E3 ligases, the degradation of Smad proteins by Nedd4 E3 ligases also regulates TGF- β signaling by removing them from the cellular environment. Smurf 2 (Smad ubiquitin regulatory factor 2) has a high affinity for the I-smad Smad 7 but is still capable of binding Smad 2 and 3 at a reduced degree (Kavsak *et al.*, 2000; Lin *et al.*, 2000). Deletion constructs of Smurf 2 identified WW2 as the binding region for Smads 2 and 3, as deletion abolished binding between these partners (Lin *et al.*, 2000). Interaction of Smurf 2 with Smad 2 resulted in ubiquitination and degradation of the smad protein during TGF- β activation. In contrast, Smad 3 expression levels were unaffected, suggesting another E3 ligase could be responsible for the majority of Smad 3 turnover. Based on the preference for Smad 3 by WWP2 isoforms regardless of TGF- β status mentioned previously, WWP2 may be the preferential controller of Smad 3 levels. This was further supported by

luciferase reporter assay where, upon TGF- β stimulation, expression of Smurf 2 resulted in a greater decrease in Smad 2 activity compared to Smad 3. Smurf 2: Smad 7 binding was found to have a slightly different outcome. Kavsak *et al.* (2000) presented evidence to support direct binding of Smurf 2 and Smad 7 and later discovered this complex was recruited to the TGF- β type I and II receptors located at the cellular membrane, rather than immediate ubiquitination of Smad 7 upon binding. The complex formation enabled Smurf 2 translocation out of the nucleus leading to down-regulation of TGF- β receptor availability, thus lowering the activity of the pathway.

Further screening assays also identified WWP1 and Nedd4-2 as Smad 7 interaction partners (Komuro *et al.*, 2004; Kuratomi *et al.*, 2005). Both WWP1 and Nedd4-2 showed a similar function as Smurf 2 in effectively binding to Smads 2, 3 and 7 as well as mediating receptor degradation using Smad 7 as an adaptor (Komuro *et al.*, 2004). However, WWP1 failed to ubiquitinate the R-smads, unlike Smurf 2, suggesting a different role for WWP1 associate in R-Smad regulation than ubiquitination for degradation. These ligases also appeared to have Smad 4 interaction capabilities via Smad 7, leading to proteasomal degradation of Smad 4, similar to the Smurf E3 ligases (Morén *et al.*, 2005). These overlapping yet individual functions suggest partial redundancy in E3 ligase activity in regulating the TGF- β pathway.

1.5.1. TGF- β signaling in prostate cancer

TGF- β signaling is known to be both tumour suppressive and tumour promoting dependent on the context of the cell. The tight regulation of this pathway is important as it has various roles in development, differentiation and cell maintenance. Understanding the deviations from controlled signaling present in non-cancerous cells is important in uncovering novel aspects of the disease which could potentially lead to better therapeutics.

The three isoforms of TGF- β (TGF- β 1, TGF- β 2 and TGF- β 3) can all bind to the same receptors and seem to share biological activities in cells. In early stages of PCa, all three TGF- β proteins are suppressive in nature by inducing cell cycle arrest and apoptosis, however as the disease progresses cells become resistant and increased TGF- β promotes angiogenesis and tumour growth (Croxford *et al.*, 2017). Over expression of TGF- β 1 in primary PCa patient samples was found to shorten survival, and increase angiogenesis and metastasis (Wikström *et al.*, 1998). Likewise, higher than normal levels of TGF- β 2 as a result of PI3K and p38 signaling also induces

tumourigenesis (Wan *et al.*, 2010). However, TGF- β 3 has shown the most potent effect by inducing invasive phenotypes by reducing cellular adhesion and increasing motility via Smad 3 signaling in PC3 cells (Walker *et al.*, 2013). Expression analysis of both mRNA and protein levels of TGF- β signaling components in prostate cancer cell lines identified both metastatic cell lines PC3 and DU145 as having high expression of TGF- β 3, potentially highlighting the role of this isoform in prostate metastasis. In contrast, TGF- β 1 was ubiquitous across all cell lines investigated, suggesting a minor role for this isoform in metastatic changes. Further investigation revealed differences in proliferation, migration and invasion between the PCa cell lines in response to TGF- β isoforms. While DU145 proliferation was inhibited by TGF- β , PC-3 cells showed no response. On the other hand, migration and invasion was increased in PC-3 cells but remained unchanged in DU145 cells. These differences may be due to the context of the cells and the use of both Smad-dependent and -independent signaling pathways.

Several reports have linked TGF- β in the induction of EMT (epithelial-to-mesenchymal transition) associated with cancer development, progression and therapeutic resistance (Khan *et al.*, 2015; Montanari *et al.*, 2017). Epithelial cells are known for their tight and adherent junctions that restrict them to a defined organization with a basement membrane. In contrast, mesenchymal cells are less rigid in structure and more migratory (Khan *et al.*, 2015). The molecular programming that controls this process relies on the cooperative action of many pathways and transcription factors. Key components are often used as markers to determine whether the morphological and cellular changes observed are related to EMT initiation. For example, E-cadherin involved in cell junctions is often down-regulated during EMT due to repression of transcription and protein phosphorylation and degradation (Tsai and Yang, 2013). Another marker is vimentin, an intermediate filament type, which increases in expression levels during EMT and PCa metastasis, as shown in the advanced metastatic PCa cell lines PC3 and DU145 (Singh *et al.*, 2003). Other proteins which are regulators of EMT are Snail and Twist, also used as markers. These proteins are known to function directly and indirectly as repressors of E-cadherin expression (Tsai and Yang, 2013), in turn inducing a more mesenchymal phenotype. TGF- β signaling has been shown to directly alter the transcription of these factors highlighting it as one of the main mechanisms for the induction of EMT, and consequently, the transition towards metastatic PCa.

The overexpression of Nedd4 E3 ligases has been identified as a major mechanism of EMT up-regulation in cancer (Soond and Chantry, 2011; Soond *et al.*, 2013). shRNA knockdown of WWP2-N appeared to redirect WWP2-FL to target Smad 7 over the R-smads 2 and 3 resulting in greater expression of the EMT marker vimentin, suggesting greater capacity for differentiation into migratory cancer cells. This was supported by analysis of other EMT markers in which the same effects were seen. Furthermore, individual WW domain expression of WW1 and WW4, present in all isoforms, resulted in inhibited Smad 3-mediated transcription. WW1 alone also reduced EMT, observed by decreased vimentin under TGF- β stimulated conditions. This indicates the potent effects these individual WW domains of WWP2 can have and why they make good targets for drug design. Similar work was carried out by Chong *et al.* (2010), using Smurf 2 WW domains and Smad 7 PY-based peptide. Further work is needed into the structural effects and the role these domains could play as biomarkers and/or therapeutic agents in PCa and potentially other cancers such as breast and melanoma, which also present with dysregulation of TGF- β and Nedd4 E3 ligases.

1.6. PTEN and PI3K/AKT signaling

Phosphoinositide 3 kinase (PI3K)-AKT signaling is an important pathway for the control of many cellular processes such as cell proliferation, differentiation, cellular metabolism and apoptosis (Vara *et al.*, 2004; Song *et al.*, 2012). Growth factor binding to its respective receptor tyrosine kinase triggers internal phosphorylation of the receptor and recruitment of phosphoinositide 3-kinase (PI3K) (Figure 1.7.1). Binding of ligand to G-protein coupled receptors can also trigger PI3K localization to the plasma membrane. PI3K converts phosphatidylinositol-4,5-biphosphate (PI-4,5-P₂, also known as PIP₂) to the triphosphate PI-3,4,5-P₃ (PIP₃) mediating further signaling via serine/threonine protein kinases PDK1 (phosphoinositide-dependent kinase 1) and AKT/PKB (previously called protein kinase B). These kinases are essential to the activation/inhibition of other proteins leading to growth and apoptosis and are often involved in cellular transformation into cancer phenotypes. Hence, understanding the regulation of these pathways is important for gaining knowledge of cancerous conditions and triggers of cancer formation, as well as for the development of novel therapeutics which are greatly required.

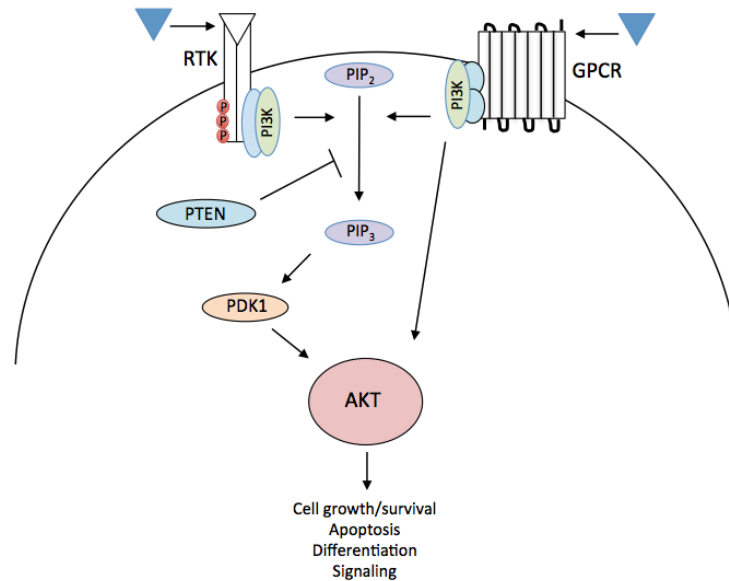


Figure 1.7.1 - Schematic representation of PI3K/AKT signaling activation

A simplified diagram showing the PI3K/AKT pathway. Binding of ligand/growth factor (blue triangle) to a receptor tyrosine kinase (RTK) or G-protein couple receptor (GPCR) (white rectangles) activates the receptors and phosphoinositide 3-kinase (PI3K) (green circle) is recruited. PI3K then converts phosphatidylinositol-4,5-biphosphate (PIP₂) into phosphatidylinositol-3,4,5-triphosphate (PIP₃) (purple circles) which then activates the kinases phosphoinositide-dependent kinase 1 (PDK1) (orange circle) and protein kinase B (AKT) (red circle) for downstream phosphorylation events that control many cellular processes such as cellular growth and survival. PTEN (phosphatase and tensin homologue deleted on chromosome 10) negatively regulates PI3K activity.

One protein known to regulate PI3K/AKT signaling is PTEN, a phosphatase and tensin homologue deleted on chromosome 10. PTEN is 403 residues in length and consists of an N-terminal PIP₂ binding domain followed by a catalytic phosphatase domain (Brito *et al.*, 2015). Next to this resides a phospholipid binding C2 domain, a C-terminal tail and finally a PDZ domain. PTEN negatively regulates PI3K/AKT signaling by catalyzing the removal of a phosphate group from PIP₃ converting it back to PIP₂. This results in reduced levels of substrate for AKT activation and a decrease in signaling activity (Song *et al.*, 2012). The role of PTEN as a negative regulator of AKT signaling has been associated with tumour suppressor activity as PTEN loss, often seen in many types of cancer (Chen and Matesic, 2007; Fang *et al.*, 2007; Guldberg *et al.*, 1997; Taylor *et al.*, 2010), leads to constitutive activation of the pathway and uncontrolled cellular growth and survival. This is further supported by diseases like Cowden's syndrome where PTEN mutation results in patients developing multiple tumours such as breast,

colorectal, kidney, prostate and melanoma (Carnero and Paramio, 2014; He *et al.*, 2012). Therefore, maintaining levels of PTEN appears to be critical for the regulation of normal cellular function.

PTEN is regulated at several stages of its production and activities. Post-translational modifications and protein-protein interactions are used to regulate localisation, interactions, enzymatic activity and turnover of this phosphatase (Brito *et al.*, 2015; Gericke *et al.*, 2006). One key modification is phosphorylation. The C-tail contains serine/threonine clusters which are targeted for phosphorylation by kinases such as casein kinase 2 (CK2) (Miller *et al.*, 2002) and GSK3 β (Al-Khouri *et al.*, 2005). The addition of phosphate groups to the tail results in a more compact conformation restricting the access of interaction partners, and also prevents membrane association. PTEN activity is predominantly associated with its lipid phosphatase domain so by preventing interaction with the plasma membrane or proteins within the membrane, this effectively inactivates PTEN. This has also been shown to occur by mutation of the PIP₂ binding domain (Walker *et al.*, 2004).

PTEN expression levels, cellular localization and activity are also regulated by ubiquitination. Both mono- and poly-ubiquitination have been reported to control PTEN activity. Wang and colleagues (2007) identified Nedd4 as a PTEN specific E3 ligase. In the presence of Nedd4, an increase in PTEN poly-ubiquitination and degradation was observed, implicating Nedd4 in PTEN turnover. Furthermore, knockdown of Nedd4 resulted in increased PTEN levels and decreased phospho-AKT signaling. However, further research contradicted Nedd4 as the major regulator of PTEN function and degradation (Chen *et al.*, 2016; Chen *et al.*, 2017; Fouladkou *et al.*, 2008). Purification of PTEN and its endogenous interaction partners failed to identify Nedd4 but did reveal WWP2, another E3 ligase in same family as Nedd4 (Maddika *et al.*, 2011). This may be a result of the cellular context; however, it has later been shown that WWP2 is a biological interaction partner capable of ubiquitinating PTEN in mouse models (Li *et al.*, 2018). Increased PTEN levels and decreased PI3K/AKT signaling was observed in mouse embryonic fibroblasts derived from WWP2 knockout mice, although PTEN mRNA was unaffected. This supports WWP2 as a physiological regulator of PTEN activity at the protein level. Further investigations into the activity of WWP2 and Nedd4 on the function of PTEN revealed that the phosphorylation status of PTEN can control the interaction and ubiquitination mediated by Nedd4 E3 ligases (Chen *et al.*, 2016). While poly-ubiquitination by Nedd4 on PTEN was unaffected by C-tail phosphorylation,

WWP2 mediated poly-ubiquitination was abolished. The phosphorylation of PTEN protects the phosphatase from ubiquitination and degradation thereby regulating steady state levels of PTEN. The targeting of the active unphosphorylated PTEN by WWP2 adds another level of tight control on the system. As for mono-ubiquitination, the addition of a single ubiquitin molecule by Nedd4 to PTEN has been shown to facilitate transport of PTEN into the nucleus (Trotman *et al.*, 2007). The localization of PTEN into the nucleus protects it from cytoplasmic targeting for degradation. Additionally, the presence of PTEN in the nucleus has been shown to antagonize AKT signaling and in turn trigger apoptosis. However, the localization of Nedd4 in the nucleus has also been shown in some cell lines, therefore PTEN mono-ubiquitination by Nedd4 may not be an entirely protective mechanism.

1.6.1. PTEN in prostate cancer

The study of PTEN loss in mouse models has demonstrated that homozygous deletion results in an embryonic lethal phenotype. However, mice with a heterozygous genotype for the PTEN allele develop several tumours in areas such as the breast and prostate (Cristofano *et al.*, 1998). This supports the identity of PTEN as a true tumour suppressor protein. With the use of conditional knockout of PTEN, Wang and colleagues (2003) showed that complete loss of PTEN not only resulted in prostate cancer development, but also significantly increased the invasion capabilities of the cancer in turn decreasing the time-frame for metastatic cancer to develop. Additionally, the combined inactivation of PTEN and Smad4 in TGF- β signaling, further enhances tumourigenesis and drives invasion towards metastatic PCa (Ding *et al.*, 2011). The loss of PTEN is also associated with reoccurrence under androgen-independent conditions and the resistance to androgen signaling based treatments such as abiraterone, an androgen synthesis inhibitor, and enzalutamide, an androgen receptor inhibitor (Bitting and Armstrong, 2013). Therefore, understanding the molecular regulation of PTEN could prove vital to cancer research into new therapies.

The regulation of PTEN by E3 ligases has been implicated in exacerbating the loss of PTEN in the progression of prostate cancer towards metastasis. The incidence of WWP2 overexpression, in particular WWP2-N, has been established in primary tumour samples of the prostate (Soond *et al.*, 2013). In addition to the identification of WWP2 as a major regulator of PTEN turnover, and that a reduction in WWP2 leads to increased PTEN levels and decreased AKT-mediated cellular survival, this implies that WWP2 is an oncoprotein that stimulates the progression of prostate cancer.

Additionally, cross-talk between PTEN function and the TGF- β pathway further suggests a role of WWP2 in mediating oncogenesis.

1.7. Inhibitors of ubiquitination

The ubiquitin-proteasome pathway has been of interest in drug discovery for many years as it is essential for the regulation of many cellular processes, and dysfunction of key members of this pathway have been linked to cancer development (Zhang and Sidhu, 2014). Targeting of the proteasome by the small molecule inhibitor Bortezomib showed some success in the treatment of myeloma and other cancers and was found to be fairly selective to cancerous over non-cancerous cells (Orlowski *et al.*, 2002; Richardson *et al.*, 2005; Zhang and Sidhu, 2014). However, by inhibiting the proteasome which acts to degrade many different cellular components, there is little specificity in the pathways affected and this results in many non-specific side effects. To this end, focus was directed towards other components of the ubiquitination cascade with the aim of increasing selectivity and effectiveness. Research into Uba1 activity revealed an inhibitor called PYR-41 that blocked the cysteine residue of Uba1 (Yang *et al.*, 2007). In doing so, a decrease in the total ubiquitinated protein level was detected suggesting that perhaps E1 inhibition may not be an ideal target either on the basis that only two E1 enzymes function in humans in the activation and transfer of ubiquitin to many E2-conjugating enzymes. Similarly, the interaction of a single E2 enzyme with many E3 ligases further suggests that inhibitors of ubiquitination should be designed for the more selective E3 ligases in order to target particular pathways.

The role of E3 ligases in cancer development and progression has led to increasing interest for drug discovery. The first group of active inhibitors against E3 ligases were the nutlins (Zhao *et al.*, 2015). These compounds were shown to be very potent with IC₅₀ values in the low micromolar and nanomolar range and specific to the RING E3 ligase Mdm2 often over expressed in human cancers. Nutlins block the ubiquitination and degradation of the important tumour suppressor protein p53, in turn triggering cell cycle arrest and/or apoptosis. This opened up the field of inhibitor discovery against E3 ligases due to the increasing evidence of over-expression in cancer. Nedd4 E3 ligases have also been promising targets in cancer therapy based on the selective binding by WW domains. This is supported by many HECT E3 ligases being over expressed in cancer cells (Komuro *et al.*, 2004; Soond and Chantry, 2011; Soond *et al.*, 2013). The high expression of these proteins promotes resistance to growth inhibition, increased proliferation and metastatic transition. However, few inhibitors of HECT E3 ligases have

currently been discovered. In 2014, the identification of bicyclic peptides and a small molecule inhibitor revealed that HECT E3 ligases could in fact be targets for novel therapeutics (Mund *et al.*, 2014). The bicyclic peptides were found to interact with the HECT domains of Smurf 2, Nedd4 and WWP1 at the E2-binding site, thereby blocking E2~Ub access and inhibiting the ubiquitination activity of these ligases. This supports the idea that peptide design could be used for reducing the activity of these ligases, however further work into this aspect of therapy design is required. As for the small molecule inhibitor in this work, Heclin was found to inhibit the ubiquitination activity of several Nedd4 E3 ligases, including WWP2, at 100 μ M, by altering the oxidation state of the cysteine residue in the HECT domain rather than blocking the E2~Ub binding site. Research by another group also identified an already established anti-depressant drug, clomipramine, as an inhibitor of the Nedd4 E3 ligase Itch (Rossi *et al.*, 2014). This was also shown to inhibit the thiolation reaction with the cysteine residue of HECT. Alternative to the catalytic cysteine being the sole site of inhibition, work by Kathman and colleagues (2015) discovered a covalent inhibitor of Nedd4-1, which blocked a non-catalytic cysteine (C⁶²⁷) residue associated with ubiquitin chain formation located in the N-lobe previously identified by Maspero *et al.* (2011) as a non-covalent ubiquitin-binding exosite. Other HECT E3 ligases such as Itch, WWP1 and WWP2 showed no inhibition by compound 3, thought to be due to the absence of this cysteine residue at the equivalent position, highlighting potential in generating selective Nedd4 family member specific inhibitors. Interestingly, while they lack a non-catalytic cysteine at the same position, these three ligases contain a cysteine residue 10 amino acids upstream from this position, providing a potential site for an equivalent processive mechanism and potential inhibitor binding site, although this has yet to be uncovered. Another research group later identified other indole-based compounds as potent inhibitors of Nedd4 (Quirit *et al.*, 2017). Indole-3-carbinol and analogues showed inhibition of Nedd4 polyubiquitination activity with IC₅₀ values ranging from 10-100s in micromolar. These inhibitors were also identified as binders of the HECT domain, further highlighting this region as a predominant site for current inhibitory activity. The different methods of discovery and mechanisms of action by these inhibitory molecules highlights the different possible ways in which the activity of these ligases can be inhibited, and further investigations could reveal a whole new group of therapeutics. In particular, there is potential for the discovery of inhibitors which act outside of the catalytic HECT domain to potentially inhibit selective interactions with substrates, although such inhibitors have yet to be developed.

1.8. Thesis Aims

Current knowledge of Nedd4 E3 ligase structure and function has been discussed in this chapter, with focus on TGF- β signaling and PTEN regulation of PI3K/AKT signaling. Both of these pathways undergo regulation by the E3 ligase WWP2. The overexpression of WWP2 has been associated with the initiation and progression of prostate cancer, thereby acting as an oncoprotein for the increased ubiquitination and degradation of Smad proteins and PTEN. Accumulating evidence is supporting this ligase as a potential target for novel therapeutics in prostate cancer. The presence of discrete binding modules within WWP2 presents an ideal drug design site, however information about the mechanisms of substrate binding and the selectivity of these domains is still lacking. Additionally, the catalytic activity located in the HECT domain provides another site of potential inhibition that could be manipulated in drug discovery to abolish the ubiquitination activity of this ligase.

The overall aim of this project is to utilize both structural and screening-based approaches for the discovery of WWP2 inhibitors as potential novel therapeutics for future use in prostate cancer treatment. In order to design peptidomimetics to interfere in the selective WW domain interactions with substrates, NMR analysis will be used to uncover the structure of WW4 in WWP2 that is a key interaction site for the inhibitory Smad 7 protein. Furthermore, while Smad protein interactions with WW4 have been investigated, little is known about Smad binding to the other WW domains. In conjunction with the significant role WWP2 plays in Smad protein and PTEN turnover in prostate cancer, interaction analysis using pull-downs will attempt to look at WWP2 isoform binding to PTEN, and NMR titrations using Smad peptides containing the PPxY motif will be carried out with WW1 and tandem domain WW1-2 with the aim of uncovering the atomic aspects of Smad protein interactions. The alternative approach of small molecule screening will also be used to identify drug-like compounds that are capable of inhibiting WWP2 enzymatic activity, with counter screening against other members of the ubiquitination pathway and Nedd4 family of E3 ligases to determine the selectivity of any hit compounds. Collectively, this work hopes to advance research into the structure-function relationship of WWP2 and to progress the development of novel treatments towards WWP2 in particular, of which inhibitors are currently lacking.

2. Material and Methods

2.1. DNA-based procedures

2.1.1. Plasmids and primers

Tables 2.1.1 and 2.1.2 list all the plasmids used throughout this project along with the source or method of synthesis. Primers used in this work are listed in Table 2.1.3.

Plasmid	Origin
pRK5-FLAG-WWP2-FL	Available in Dr Andrew Chantry's lab
pRK5-FLAG-WWP2-N	Available in Dr Andrew Chantry's lab
pRK5-FLAG-WWP2-C	Available in Dr Andrew Chantry's lab
pRK5-HA-WWP2-FL	Available in Dr Andrew Chantry's lab
pRK5-HA-WWP2-WW1-2	Available in Dr Andrew Chantry's lab
pRK5-HA-WWP2-WW3-4	Available in Dr Andrew Chantry's lab
pRK5-HA-Smad 2	Available in Dr Andrew Chantry's lab
pRK5-Smad 3-N-GFP	Available in Dr Andrew Chantry's lab
pSKDuet01-GB1-WW1-2	Available in Dr Andrew Chantry's lab
pOPINf-WWP1	Available in Dr Andrew Chantry's lab
pGEX2T-PTEN	Gifted by William Sellers (Addgene plasmid #10784 – 803)
pGEX2T-PTEN 1-274	Gifted by William Sellers (Addgene plasmid #10741 – 821)
pGEX2T-PTEN Δ 274-342	Gifted by William Sellers (Addgene plasmid #10739 – 824)
pRK5-HA-Ubiquitin	Gifted by Ted Dawson (Addgene plasmid #17608)
pET3a-hUBA1-CtermHis	Gifted by Titia Sixma (Addgene plasmid #63571)
pET3a-Ubch7-CtermHis	Gifted by Prof Martin Scheffner
pGEX4T-1-Ubch7a	Gifted by Dr Arthur L Haas lab
pGEX4T-1-hE1 (Uba1)	Gifted by Dr Arthur L Haas lab
pGEX-5X3-Nedd4-FL	Gifted by Shiaw-Yih Lin (Addgene plasmid #45043)

Table 2.1.1 - Summary of plasmid names and sources.

Plasmid	Cloning technique
pSKDuet01-GB1-WW1	In-fusion
pGEX2T-PTEN 1-100	QuikChange mutagenesis
pGEX2T-PTEN 1-185	QuikChange mutagenesis
pGEX2T-PTEN 1-350	QuikChange mutagenesis
pGEX2T-PTEN 1-400	QuikChange mutagenesis
pET21d-SUMO-Smad 2	In-fusion
pET21d-SUMO-Smad 3	In-fusion
pGEX2T-WWP2-C	In-fusion
pGEX2T-WWP2-FL	In-fusion
pGEX2T-WWP2-C ^{C399A}	QuikChange mutagenesis
pGEX2T-WWP2-FL ^{C838A}	QuikChange mutagenesis

Table 2.1.2 - Summary of plasmids cloned during this work and the method used.

PCR product	Vector	Primers 5' – 3'
WW1	pSKDuet-01(GB1)	F - cgtaacggaaggatcccaggccccgacgctct R - atgcggccgcaagctttaccgctcccagggtggtggtctt
PTEN 1-100a/b	pGEX2T	caagatcttcacaaaagggtcagataagttctagctgtggtggg cccaccacagctagaacttatctgaccctttgtgaagatcttg
PTEN 1-185a/b	pGEX2T	gaaacaacagtgccactggtctataatctcaatgattctttaacaggtagctataa- taa ttattatagctacctgttaaagaatcattgagattatagaccagtggcactgtgtttc
PTEN 1-350a/b	pGEX2T	tgacggctcctctcatgtttgtgaagtacagcttcacc ggtgaagctgtacttcacaaaaacatgagaggagccgtca
PTEN 1-400a/b	pGEX2T	gatgaattctcagacttttcaaattgtgtatgctgatcttcatcaaaagggtc gaacctttgatgaagatcagcatcacaaaattgaaaagtct-gagaattcatc
Smad 2	pGEX4T	F - gctagcatgtcgtccatcttgcca R - gaattcaaatgcagggtctgagta
Sm2 p1	pET21d	F - gaacagattggaggtccacagagtaattatattcc R - gctcgaattcggatccttattctccatcttcaactgat
Sm3 p1	pET21d	F - gaacagattggaggtcccagagcaatattccaga R - gctcgaattcggatccttattctccatcttcaactcaggt
Sm2 p2	pET21d	F - gaacagattggaggtattgagccacagagtaatt
Sm3 p2	pET21d	F - gaacagattggaggtatcgagccccagagcaatatt
WWP2-C	pGEX2T	F - gttccgctggatccatgatccaggaaccagct R - agtcacgatgaattcttactcctgtccaaagcc
WWP2- ^{C399A} a/b	pGEX2T	tccagacggtgaaggcgggtgtggcttctggg cccagaagccacaccgccttcaaccgtctgga
WWP2-FL	pGEX2T	F - gttccgctggatccatggcatctgccagctct R - agtcacgatgaattcttactcctgtccaaagcc
WWP2-FL ^{C838A} a/b	pGEX2T	Same mutagenesis primers as WWP2- ^{C399A} a/b

Table 2.1.3 - Summary of primers used in plasmid cloning and mutagenesis.

Sm2 – Smad 2; Sm3 – Smad 3; p1 – peptide 1; p2 – peptide 2.

2.1.2. Phusion PCR

Phusion polymerase reactions were set up as described in Table 2.1.4 per the manufacturer's protocol (Thermo Fisher #F-530S). Cycling conditions used are described in Table 2.1.5. The reaction products were analysed by 1% agarose gel electrophoresis run at 100 V using SYBR safe (Invitrogen) for 1 hour (hr) or until the dye front was ~0.5 cm from the gel end. Gels were imaged using BIORAD Geldoc+ and Imagelab (BIORAD).

Component	20 μ L rxn	100 μ L rxn	Final conc.
Sterile nuclease free H ₂ O	Add to 20 μ L	Add to 100 μ L	-
5X Phusion HF buffer	4 μ L	20 μ L	1 X
10 mM dNTPs	0.4 μ L	2 μ L	200 μ M of each
Forward primer (10 μ M)	1 μ L	5 μ L	0.5 μ M
Reverse primer (10 μ M)	1 μ L	5 μ L	0.5 μ M
Template DNA	0.25 μ L	1.25 μ L	-
Phusion DNA polymerase	0.2 μ L	1 μ L	0.02 U/ μ L

Table 2.1.4 - Components used for both small and large-scale phusion reactions (rxn).

HF - high fidelity; dNTPs – deoxyribonucleic acid triphosphates; conc – concentration; U - units.

Cycle step	Temperature ($^{\circ}$ C)	Time	No. of cycles
Initial denaturation	98	30 s	1
Denaturation	98	7.5 s	1
Annealing	Based on primer T _m	20 s	30
Extension	72	15 s/kb	1
Final extension	72	7.5 min	1
	4	Hold	

Table 2.1.5 - Cycling conditions used for phusion reactions.

T_m – melting temperature; kb – kilobases; s - seconds; min - minutes.

2.1.3. DNA gel extraction

QIAgen QIAquick gel extraction kit was used according to the manufactures protocol. DNA was eluted in 50 μ L elution buffer provided and stored on ice or at -20 $^{\circ}$ C until required.

2.1.4. In-fusion

In-fusion reactions were set up as described in Table 2.1.6 Reactions were mixed and incubated at 50 $^{\circ}$ C for 15 min, then placed on ice according to the In-fusion HD cloning kit manual (Takara Bio Clontech). 2.5 μ L reaction mix was transformed into Stellar

competent cells using the protocol provided by the manufacturer with the cells (Clontech, Takara Bio).

Component	10 μ L rxn
Sterile nuclease free H ₂ O	Add to 10 μ L
Purified PCR product	50 – 100 ng
Linear vector	50 – 100 ng
5x In-fusion HD enzyme premix	2 μ L

Table 2.1.6 - Components used for a single reaction (rxn) mix.

2.1.5. Restriction digest

Under sterile conditions, all components described in Table 2.1.7 were mixed and incubated at 37 °C for 1-4 hrs. 6x agarose loading dye (Invitrogen) was added to each sample and analysed by 1 % agarose gel electrophoresis.

Component	20 μ L rxn
Sterile nuclease free H ₂ O	Add to 20 μ L
10 x Enzyme buffer	2 μ L
BSA (1 mg/mL)	2 μ L
Plasmid DNA	1-2 μ L
Enzyme(s) - total	1 μ L

Table 2.1.7 - Restriction digest reaction components per 20 μ L reaction (rxn).

BSA – bovine serum albumin.

2.1.6. QuikChange mutagenesis

Samples were prepared according to the protocols provided with the QuikChange mutagenesis kit (Agilent Technologies). Table 2.1.8 summarizes the components used in each reaction, and primers are described in Table 2.1.3. The reaction mix was transformed into XL-10 Gold cells as described in the manufacturer's protocol provided with the kit.

Component	50 μ L rxn
dH ₂ O	Add to 50 μ L
10 x reaction buffer	5 μ L
Template DNA	10 ng
Oligo primer a	125 ng
Oligo primer b	125 ng
dNTP mix	1 μ L
QuikSolution	3 μ L

Table 2.1.8 - Component volumes used for a single mutagenesis reaction (rxn).

2.1.7. Heat shock transformation

Preparations of lysogeny broth (LB; per litre: 25 g high salt LB granules (Thermofisher; 10g peptone, 5 g yeast extract and 10 g NaCl)) and LB-agar (LB with 3 % agar) were made and autoclaved in advance. 50-100 μ L competent cells were aliquot into pre-chilled micro-centrifuge tubes. 10 ng – 1 μ g plasmid DNA/cloning reaction was added and gently mixed. After 30 min incubation on ice, samples were heat shocked for 45 seconds at 42 °C using a water bath followed by 10 min incubation on ice. 0.3 mL LB was added and the tubes incubated for 45-60 min at 37 °C and 180 rpm. 100-200 μ L transformation mixture was spread onto LB agar plates containing the appropriate antibiotics (final working concentrations: kanamycin – 50 μ g/mL; ampicillin – 100 μ g/mL; chloramphenicol – 25 μ g/mL). Plates were then incubated overnight at 37 °C.

2.1.8. Plasmid purification

Colonies were picked and used to inoculate antibiotic-containing LB and grown overnight at 37 °C. QIAgen spin miniprep kit, QIAgen plasmid plus midi kit and Nucleobond xtra maxi kit were used, dependent on DNA yield required, using the reagents and protocols provided in the kits. DNA was eluted in elution buffer provided in the kit or TE buffer (10 mM Tris-Cl pH 8, 1 mM EDTA pH 8). DNA concentration was determined using Thermo Scientific NanoDrop spectrophotometer. Sequencing was carried out to confirm plasmid DNA was correct, using Eurofins Genetics value read tube service.

2.2. Protein-based procedures

2.2.1. Recombinant protein expression - unlabelled

Plasmids were transformed into an *E.coli* BL21-derived cell line and spread onto LB agar plates, and grown overnight at 37 °C. Under aseptic conditions, a single colony was picked and used to inoculate 10 - 20 mL of antibiotic-containing LB, then incubated overnight at 37 °C and 180 rpm. Overnight cultures were used to inoculate 50 mL - 2 L antibiotic-containing LB at 1 in 50 dilutions and grown at 37 °C and 180 rpm until an OD₆₀₀ of 0.6 - 0.8 was reached. 1M IPTG (isopropyl β -D-1-thiogalactopyranoside) stock was used to induce the cultures at the concentrations and growth conditions described in Table 2.2.1.

Plasmid	Cell line	Antibiotic	IPTG conc. (mM)	Temp. (°C)	Time
pGEX2T-PTEN	BL21 (DE3)/ Rosetta2 pLysS	Amp/Amp & Cam	0.75	30/16	O/N
pGEX2T-PTEN 1-274	Rosetta2 pLysS	Amp & Cam	0.75	30	O/N
pGEX2T-PTEN Δ 274-342	Rosetta2 pLysS	Amp & Cam	0.75	30	O/N
pGEX2T-PTEN 1-100	Rosetta2 pLysS	Amp & Cam	0.75	33	O/N
pGEX2T-PTEN1-185	Rosetta2 pLysS	Amp & Cam	0.75	30	O/N
pGEX2T-PTEN1-350	Rosetta2 pLysS	Amp & Cam	0.75	30	O/N
pET21d-SUMO- Smad 2	Codon Plus RP	Amp & Cam	0.75	33	O/N
pET21d-SUMO- Smad 3	Codon Plus RP	Amp & Cam	0.75	33	O/N
pGEX2T-WWP2- C/C ^{C399A}	Rosetta2 pLysS	Amp & Cam	0.75	33	O/N
pGEX2T-WWP2- FL/FL ^{C838A}	Rosetta2 pLysS	Amp & Cam	0.75/0.5	30/16	O/N
pET3a-UbcH7	Codon Plus RP	Amp & Cam	1.0	25	4hrs
pET3a-hUBA1	BL21 (DE3)	Amp	1.0	25	O/N
pGEX4T-1-UbcH7a	BL21 (DE3)	Amp	0.4	20	O/N
pGEX4T-1-hE1 (Uba1)	BL21 (DE3)	Amp	1.0	25	O/N
pOPINF-WWP1-FL	Codon Plus RP	Amp & Cam	0.5	16	O/N

Table 2.2.1 - Protein expression conditions used for different plasmids.

Amp – ampicillin; Cam – chloramphenicol; IPTG -isopropyl β -D-1-thiogalactopyranoside; O/N – overnight.

2.2.2. Recombinant protein expression - labelled

In advance, unlabelled M9 salts (per litre: 6 g Na₂HPO₄, 3 g KH₂PO₄, 0.5 g NaCl, 1 g NH₄Cl, pH7-7.4) and ¹⁵N-labelled M9 salts (per litre: 6 g Na₂HPO₄, 3 g KH₂PO₄, 0.5 g NaCl, 1 g ¹⁵N-NH₄Cl, pH7-7.4) were made and autoclaved. Heat shock transformation of pSKDuet01-GB1-WW1 and pSKDuet01-GB1-WW1-2 into BL21 (DE3) Star was carried out as described in section 2.1.7. A single colony was picked from the LB-kanamycin agar plates to inoculate 10 mL LB-kanamycin. The starter culture was incubated overnight at 37 °C and 180 rpm. 8 mL of overnight culture was used to inoculate 100 mL of unlabelled minimal media plus supplements (final concentrations:

2 mM MgSO₄, 10 µM CaCl₂, 0.1 mM FeSO₄ (from a stock made fresh), 0.2 % (w/v) glucose, 0.1 % vitamins solution (SIGMA M6895), 0.1 % micronutrients (1000 x stock: 3 µM (NH₄)₂MoO₄, 4 µM H₃BO₃, 30 µM CaCl₂, 10 µM CuSO₄, 80 µM MnCl₂, 10 µM ZnSO₄) and left to grow at 37 °C, 180 rpm, until saturation. The starter culture was centrifuged at 4300 x g for 20 min at 22 °C and the supernatant discarded. The pellet was resuspended in 10 mL ¹⁵N-labelled minimal media, plus supplements at the same final concentrations as above, and added to each 500 mL ¹⁵N-labelled minimal media. Double-labelled ¹³C-¹⁵N media was made and used in the same way except 2 g/L ¹³C-glucose was used. Large-scale cultures were incubated at 37 °C until an OD₆₀₀ 0.6—0.8 was reached. Cultures were induced with 1 M IPTG at a final concentration of 0.75 mM and incubated at 30 °C and 180 rpm overnight. Cultures were centrifuged at 4300 x g 4 °C for 20 min. The supernatants were discarded, and pellets stored at -80 °C until required.

2.2.3. Cell lysate preparation

Cell pellets were resuspended in appropriate buffers and sonicated on ice at 70-80 % amp for pulses of 2 sec on 1 sec off for a total of 2 min (Sonics Vibracell VCX 130). This was repeated 3-6 times with incubations on ice in between to prevent over-heating of the suspensions. Cell suspensions were then centrifuged either at 18 krpm 4 °C for 45 min (Beckman, JA-20 rotor) or at 4300 x g 4 °C for 30 min (Heraeus Megafuge 1.0R). Supernatants were decanted into a centrifuge tube for further purification and the pellet discarded. Samples were taken of the whole, insoluble and soluble fractions for SDS-PAGE analysis.

2.2.4. Protein purification methods

All buffers used in the following methods were filtered with either a 0.44 µm or 0.22 µm sterile filter.

2.2.4.1. Nickel NTA

Cell lysates were prepared as described in section 2.2.3 using Ni-NTA binding buffer (20 mM Na₂HPO₄, 0.5 M NaCl, 30 mM imidazole pH 7.4) and placed on ice until required. HisTrap FF 1mL and 5 mL columns (GE Healthcare) were used, according to the protocol provided with the columns, with 300 mM imidazole in the elution buffer (20 mM Na₂HPO₄, 0.5 M NaCl, 300 mM imidazole). The flow-through during sample loading, wash and manual elution were collected. Aliquots were taken for SDS-PAGE analysis. For the purification of His-GB1-WW1 and His-GB1-WW1-2, the ÄKTA-FPLC purification system was used for elution using a linear gradient of imidazole from 30 mM

to 300 mM at 1-5 mL/min flow-rate. 2-5 mL fractions were collected, and aliquots taken for SDS-PAGE analysis.

2.2.4.2. Protein tag cleavage

His-GB1-WW1 and His-GB1-WW1-2 were dialysed overnight at room temperature into Tris-Cl buffer (25 mM Tris-Cl pH 8, 0.5 M NaCl) using 7 kDa MWCO snakeskin dialysis tubing (Thermo Fisher). His(6x)-tag was cleaved off using thrombin (Thermo Fisher) at 2 U/mg of protein, added to the dialysis sample. The samples were transferred to centrifuge tubes to continue digesting until 90-100 % digestion was achieved as determined by SDS-PAGE analysis. His-SUMO-Smad 2 and Smad 3 peptides were cleaved using His-SUMO protease (Invitrogen) at 1 U/mg or His-ULP-1 at 0.5:60 in Tris-Cl buffer as described above. Samples were analysed by SDS-PAGE. GST-tagged Ubch7 and GST-WWP2-FL were digested via on-column digestion overnight at 4 °C using a GSTrap FF 1 mL column (GE Healthcare) loaded with PBS (10 mM Na₂HPO₄, 2 mM KH₂PO₄, 2.7 mM KCl, 137 mM NaCl pH 7.4) containing 10 U/mL thrombin.

2.2.4.3. Benzamidine

Benzamidine resin (GE healthcare) was used for the removal of thrombin from GB1-WW1, GB1-WW1-2 and Ubch7 digestion samples. A gravity flow column was made with a 2 cm bed of resin. The column was washed with 5 column volumes dH₂O and equilibrated with 5 column volumes PBS. The digest sample salt concentration was increased from 137 mM NaCl to 0.5 M then loaded onto the column. The flow-through was collected and the column washed with 5 column volumes high-salt PBS (10 mM Na₂HPO₄, 2 mM KH₂PO₄, 2.7 mM KCl, 0.5 M NaCl pH 7.4). After the wash was collected, any bound proteins were eluted using elution buffer (50 mM glycine pH 3.0) into a centrifuge tube containing Tris-Cl pH 7.4 to prevent pH induced protein denaturation in case protein of interest was bound to the column. Samples were taken for SDS-PAGE analysis.

2.2.4.4. Size exclusion chromatography

A HiLoad 16/200 Sepharose S75 gel filtration column (GE Healthcare) was attached to the ÄKTA-FPLC system and washed with 2 column volumes of dH₂O. The column was equilibrated with gel filtration buffer (50 mM Tris, 150 mM NaCl pH 7.5). After concentration of the protein sample to 2 mL, using a 5 kDa cut-off Vivaspinn 20 concentrator at 4300 xg 4 °C, the sample was loaded into the ÄKTA sample loop and loaded onto the column. A total of 120 mL buffer was run through the column at

1 mL/min and 2 mL fractions collected. Aliquots of alternate fractions were analysed by SDS-PAGE to determine the cleanest fractions to be pooled and dialysed.

2.2.4.5. Glutathione-GST

Cell lysates were prepared as described in section 2.2.3 using PBS or PBS with protease inhibitors (Invitrogen). GST-tagged proteins were either purified using GSTrap FF 1 mL columns (GE healthcare) or by batch purification using glutathione sepharose 4B resin (GE healthcare). Protocols provided by the manufacturer were followed. Samples were taken of the flow-through, wash and elution for the column method and an aliquot was taken of the beads from batch purification for SDS-PAGE analysis.

2.2.5. GST-pull down assay

PTEN coated beads were made as described in section 2.2.4.5 using 200 μ L of 50 % glutathione bead slurry in PBS, with the PTEN proteins retained on the beads. ~5-100 μ L of the slurry was used per reaction, based on protein expression and binding as analysed by SDS-PAGE. 50 μ L cell lysate containing HA-tagged WWP2-FL (full length), WWP2-N (N-terminal C2 and WW1), WWP2-C (C-terminal HECT and WW4), WW tandem domain 1-2 or WW 3-4 was added, and the mixture incubated for 1 hr at 4 °C. Samples were centrifuged at 5000 rpm for 30 secs – 1min to pellet the beads. The supernatant was removed, and the beads washed 4-6 times with PBST (PBS containing 0.1 % Tween). 2x Laemmli (125 mM Tris, 4 % (w/v) SDS, 20 % (v/v) glycerol, 0.01 % (w/v) bromophenol blue 100 mM DTT pH 6.8) was added and the samples heated at 100 °C for 3 min to elute and denature the proteins. Samples were analysed by SDS-PAGE and Western blot.

2.2.6. SDS-PAGE

2.2.6.1. Tris-glycine gels

Tris-glycine resolving and stacking gels were made using the components described in Table 2.2.2. APS and TEMED were added last, gently mixed in and the solution poured between glass plates to produce a gel of 1.5 mm thickness with 8 or 15 wells. Gels were stored in SDS-PAGE running buffer (25 mM Tris base, 190 mM glycine, 0.1 % SDS, pH 8.3) at 4 °C until required, for a maximum of 2 weeks. Equal volume of 2x Laemmli buffer was added to each sample and heated for 3 min at 100 °C. Samples were analysed by 10-15 % SDS-PAGE, set up as per manufacturer's instructions for BIORAD mini-protean tetra system in SDS-PAGE running buffer for 60 min at 180 V or until the dye front reached the bottom. The gel was stained with Instant Blue (Expedeon) for a minimum of an hour and imaged using epiwhite light imager. Broad

range SDS-PAGE molecular weight standard (BIO-RAD, Catalogue no. 161-0317) was used for molecular weight analysis.

Component	Resolving gel		Stacking gel 6 %
	10 %	15 %	
dH ₂ O	7.7 mL	5.7 mL	5.8 mL
37.5:1 40 % (w/v) Acrylamide	4 mL	6 mL	1.5 mL
1.5 M Tris pH 8.8	4 mL	4 mL	-
0.5 M Tris pH 6.8	-	-	2.5 mL
10 % (w/v) SDS	160 µL	160 µL	100 µL
10 % (w/v) APS	160 µL	160 µL	100 µL
TEMED	16 µL	16 µL	10 µL

Table 2.2.2 - Gel components and volumes per 2 gels at 1.5 mm thickness.

SDS – sodium dodecyl sulphate; APS – ammonium persulphate; TEMED – tetramethylethylenediamine.

2.2.6.2. Tricine gels

Tricine gels were used for resolution of low molecular weight proteins such as peptides. Gels were made using the same protocol as described in section 2.2.6.1 using the components described in Table 2.2.3 at both 1.5 mm and 0.75 mm thickness. Gels were run as described in the manufacturer's instructions for BIO-RAD mini-protean tetra system (BIO-RAD) using upper cathode buffer (1 M Tris, 1 M Tricine and 1 % SDS pH 8.25) and lower anode buffer (2.1 M Tris pH 8.9). Gels were stained using silver staining.

Component	Resolving gel	Stacking gel 6 %
dH ₂ O	3.1 mL	6.72 mL
19:1 40 % (w/v) Bis-Acrylamide	5 mL	1.06 mL
Gel buffer (3 M Tris-Cl pH 8.45, 0.3 % SDS)	5 mL	2.5 mL
70 % Glycerol	2 mL	-
10 % (w/v) APS	66.5 µL	80 µL
TEMED	6.6 µL	8 µL

Table 2.2.3 - Components and volumes used for making 2 tricine gels at 0.75 mm thickness.

APS – ammonium persulphate; TEMED – tetramethylethylenediamine.

2.2.6.3. Silver staining

Tricine gels were fixed for 30 min using a solution of 40 % methanol and 10 % acetic acid. The gels were then incubated for 10 min in 5 % methanol followed by 10 min in DTT solution (0.4 μ M DTT in dH₂O). Gels were then very briefly coated in dH₂O and the water drained off. 0.1 % AgNO₃ solution was added and incubated for 10 min. Gels were washed again in dH₂O for 5-10 sec three times before developing solution wash added (7.5 g Na₂CO₃, 250 mL dH₂O, 125 μ L formaldehyde). The reaction was stopped using solid citric acid when the peptide bands appeared.

2.2.7. Western blotting

Samples were prepared and separated using a 10/15 % polyacrylamide gel as described above. The proteins were transferred onto 0.2 μ m nitrocellulose membrane (BIORAD) using the Abcam protocol for a wet transfer. The transfer was run for 60 min at 100 V in transfer buffer (25 mM Tris base, 190 mM glycine, 20 % methanol). 0.1 % Ponceau S (0.1 % (w/v) in 5 % acetic acid) was used to check for protein presence and the blots then washed with TBST (Tris buffered saline-tween) (140 mM NaCl, 3 mM KCl, 25 mM Tris pH 7.4, 0.1 % Tween-20). The membranes were blocked with blocking buffer (3 % BSA (bovine serum albumin), 0.01 % sodium azide, TBST) for 1 hr at 4 °C with agitation before incubation in primary antibody solution, described in Table 2.2.4, overnight at 4 °C with agitation. After incubation and three TBST washes, secondary antibody diluted 1 in 10,000 in TBST was incubated with the blots for 1 hr at room temperature with agitation. After three TBST washes, the blot was developed by ECL chemiluminescence (Amersham ECL primer Western blotting detection reagent, GE Healthcare) according to the kit protocol. Amersham hyperfilm ECL high performance ECL film (GE Healthcare) was used and hand developed using Kodak film developer and fixer in a dark room. Conjugated primary antibodies were only incubated at room temperature for 1 hr before washing and development.

Antibody	Dilution and buffer	Source
Anti-GST (mouse)	1 in 2000	Thermo Fisher 8-326
Anti-HA (rat)	1 in 1000	Sigma Clone3F10
Anti-Flag (mouse)	1 in 10,000	Sigma F3165
Anti-6xHis-HRP	1 in 20,000	Abcam Ab1187
Anti-streptavidin-HRP	1 in 10,000	Thermo Fisher N100
Anti-Flag-HRP	1 in 10,000	Sigma A8592
Anti-mouse	1 in 10,000	Sigma A4416
Anti-rat	1 in 10,000	Sigma A5795

Table 2.2.4 - Summary of antibodies, dilutions and source used.

GST – glutathione S transferase; HA – human influenza hemagglutinin-derived tag; HRP – horse radish peroxidase.

2.2.8. Glutathione plate coating

Clear 96 well plates were incubated overnight at 4 °C with 50 µL of 2% (w/v) haemoglobin in 50 mM Na₂CO₃. Plates were then washed 4-6 times with PBS, incubated for 1 hr at room temperature in 50 µL 0.1 mM SSMPB (sulpho-succinimidyl 4-(p-maleimidophenyl) butyrate) (Thermo Fisher) in PBS, washed a further 4 times with PBS then incubated overnight in 10 mM reduced glutathione solution (10 mM reduced glutathione, 10 mM Na₂HPO₄, 0.15 M NaCl, 1 mM EDTA pH 6.7). Plates were washed 4 times and then stored in PBST at 4 °C.

2.2.9. In vitro auto-ubiquitination assays

All incubations were carried out at room temperature unless otherwise stated.

2.2.9.1. Glutathione bead auto-ubiquitination assay

100 µL 50 % glutathione sepharose 4B resin was washed with PBS to remove storage ethanol then incubated on a tube rotator (Labinco SH010) for 1 hr at 4 °C with 5 mL cell lysate containing either GST-WWP2-C or GST-WWP2-C^{C399A}. Tubes were centrifuged at 600 x g 4 °C for 1 min to pellet the beads. The supernatant was discarded, and beads washed 3 times with PBST. 50 ng Uba1 (E1), 240 ng Ubch7 (E2) and 5 µg biotinylated-ubiquitin was added to 10 µL bead samples to a total volume of 60 µL in ubiquitination buffer (50 mM Tris pH 7.5, 2 mM ATP, 5 mM MgCl₂, 2 mM DTT). For heclin samples, concentrations ranged from 6 – 100 µM with a total volume of 60 µL. Beads were incubated for 1 hr then washed using PBST. Samples were eluted from the beads using 2x Laemlli buffer and heating at 100 °C for 3 min. Samples were then analysed by SDS-PAGE and Western blotting using anti-streptavidin-HRP at a 1 in 10,000 dilution in TBST.

2.2.9.2. GST-WWP2 auto-ubiquitination assay

Storage PBST buffer was removed and plates tapped dry prior to assay set up. After each incubation plates were washed 3 times with PBST and tapped dry. 50 μ L cell lysate containing GST-tagged WWP2 was incubated on glutathione coated plates for 1 hr. During plate incubation, 5-20 ng/well His-E1, 120-200 ng/well His-E2, 100 ng/well Flag-ubiquitin and 1.25 mM ATP were mixed together in ubiquitination buffer (25 mM Tris pH 8, 100 mM NaCl, 4 mM $MgCl_2$, 50 μ M DTT) and incubated for 45 min. 10 μ M 0.1 % DMSO (0.5-100 μ M 1 % DMSO in the dose experiments) NCI diversity set V/natural products IV/oncology set VII compound (NIH/NCI DTP, <http://dtp.cancer.gov>) and 10/18 μ L reaction mix to a total of 20 μ L was added to each well. At least 3 wells per plate were left without reaction mix as 0 % activity controls (E3 only). 0.1/1 % DMSO was added to at least 3 wells per plate for 100 % activity control reactions. 3 wells of buffer were also used per plate to remove buffer effects. Plates were then incubated for 2 hrs and washed as above. 1 in 10,000 diluted anti-FLAG-HRP conjugate antibody solution (Sigma A8592) was added to all wells and incubated for 1 hr. Levels of anti-FLAG antibody were quantified using TMB ELISA solution (Invitrogen) which was stopped using 1 M HCl and the OD at 450 nm measured using a Hidex Sense microplate reader.

2.2.9.3. His-E2 assay

Reaction mixes containing 2.2 μ g/well GST-E1, 200 ng/well His-E2, 100 ng/well Flag-ubiquitin, 1.25 mM ATP, 10 μ M compound or 1 % DMSO and ubiquitination buffer were set up. Mixes were incubated for 1 hr. Clear 96 well nickel-coated plates (Pierce, Thermo Fisher) were washed with PBST and tapped dry before 20 μ L reaction mix was aliquoted into wells and incubated for 1 hr. Positive control wells contained DMSO-reaction mix and negative control wells contained only His-E2 and buffer. After incubation, plates were washed 3 times with PBST and incubated with anti-Flag-HRP conjugate solution for 1 hr. Plates were washed 3 times with PBST and tapped dry. Quantification was carried as described in section 2.2.9.2.

2.2.9.4. GST-E1 assays

These assays were set up and run similar to that described in section 2.2.9.2 however, the cell lysate contained GST-E1 protein and reaction mixes consisted of 100 ng/well Flag-ubiquitin, 1.25 mM ATP and 10 μ M 0.1% DMSO compounds. Reaction mix was incubated on the GST-E1-coated plates for 1 hr before washing with PBST. Plates were processed as described in section 2.2.9.2.

2.2.9.5. His-E1 ubiquitination assay

Reactions were set up containing 0.9 µg His-E1, 1 µg ubiquitin, 2.5 mM ATP in ubiquitination buffer as described in section 2.2.9.2, with 0.1 % DMSO, 10 µM compound in 0.1 % DMSO, or 1 % β-mercaptoethanol as required to a total volume of 20 µL. All reactions were incubated for 1 hr and analysed by SDS-PAGE and Instant Blue staining.

2.2.9.6. GST-Nedd4 counter assay

GST-Nedd4 auto-ubiquitination reactions were set up and run in the same manner as GST-WWP2 assays as described in section 2.2.9.2, using lysate from cells expressing GST-Nedd4 with compounds at 10 µM 0.1 % DMSO. All other components and steps were the same.

2.2.9.7. His-WWP1 counter assay

Cell lysate expressing His-WWP1 was incubated on clear 96 well plates nickel-coated (Pierce, Thermo Fisher) for 1 hr. Plates were washed 3 times with PBST containing 15 mM imidazole. Plate blocking was carried out for 1 hr using 0.5 % BSA in PBST. After washing, 18 µL reaction mixes were added to the appropriate wells, containing 2.7 µg/well GST-E1, 11 µg/well untagged Ubch7, 100 ng/well flag-ubiquitin and 1.25 mM ATP as required after being incubated for 45 min prior to addition. 10 µM compound at 0.1 % DMSO were used. Plates were incubated for 2 hrs at room temperature and washed with PBST containing 15 mM imidazole. All other steps were the same as described in section 2.2.9.2.

2.2.9.8. GST-PTEN substrate ubiquitination assay

Reactions were set up containing the following components as required: 100 nM His-E1, 1 µM His-E2, 2 µM GST-WWP2-FL, 10 µM GST-PTEN, 250 µM Flag-ubiquitin and 5 mM ATP in ubiquitination buffer (40 mM Tris-Cl pH 7.5, 50 mM NaCl, 2 mM DTT, 5 mM MgCl₂) to a total volume of 60 µL. All samples were incubated for 1 hr at room temperature. Samples were analysed by SDS-PAGE and Western blotting against GST, Flag and His-tags.

2.2.10. Statistical analysis

All graphs and statistics were generated using GraphPad Prism 7.0c for Mac OS X (GraphPad Software, La Jolla California, USA; www.graphpad.com). For initial compound screening, Z score was used to determine the quality of the plate assays. For comparison to controls, a one-way or two-way ANOVA with Tukey's, Dunnet's or Sidek multiple comparisons were used.

2.3. Cell-based procedures

All cells were grown at 37 °C and 5 % CO₂ unless otherwise stated.

2.3.1. Cell line preparation

Cells were removed from the -80 °C freezer and immediately thawed using a 37 °C water bath with gentle swirling. When only a little ice was remaining, vials were transferred into the sterile lamina flow hood (Herasafe) and the cells added to 5 mL media, as described in Table 2.3.1. Cell suspensions were centrifuged at 600-1200 x g 22 °C for 5 min to pellet the cells. The supernatant was discarded, and the pellet resuspended in 1 mL fresh media. Cells were transferred to a T25/T75 sterile cell culture flask (Nunclon Delta single vent tilt flasks, Thermo Fisher) and left to grow in an incubator.

Cell line	Cell Type	Media composition
HEK 293A	Human embryonic kidney cells	Dulbecco's Modified Eagle's Medium (DMEM), 10 % Foetal calf serum (FCS), 1% penicillin-streptomycin (P/S), 1 % non-essential amino acids, 1 % L-glutamine
DU145 PC-3	Human metastatic prostate cancer cells	Roswell Park Memorial Institute (RPMI) 1640 medium, 10 % FCS, 1 % L-glutamine, 1 % P/S

Table 2.3.1 - Mammalian cell lines and culture media used in this research.

2.3.2. Cell maintenance

Cells were grown up to 70-80 % confluency. Media was removed, and cells washed with DPBS (Dulbecco's PBS) (Gibco, Thermo Fisher). 2-5 mL TrypLE Express (Gibco, Thermo Fisher) was added and the flask incubated for 3 min at 37 °C. The contents of the flask were transferred to a centrifuge tube containing 2-5 mL media and centrifuged at 1200 x g 22 °C for 5 min. The supernatant was discarded, and the cell pellet resuspended in 1 mL media. A proportion of the 1 mL cell suspension was added to 5/10 mL fresh media in a T25/T75 flask at dilutions ranging from 1 in 10 to 1 in 3. Flasks were transferred to an incubator and the cells left to grow.

2.3.3. 6-24 well plate preparation

Cells were prepared as described in section 2.3.2, however the pellet was resuspended in 5 mL/well per 6 well plate and 1 mL/well per 24 well plate (Nunclon Delta 6- and 24-well, Thermo Fisher). Plates were incubated to ~80 % confluency.

2.3.4. Scratch assays

24 well plates were grown until a monolayer had formed in the wells. A P2 (Gilson) pipette tip was used to create a scratch. After media was removed, the wells were washed with DPBS. 5 % FBS RPMI 1640 (and supplements described in Table 2.3.1) containing 0.1 % DMSO or 10 μ M compound in 0.1 % DMSO was added to the appropriate wells. Blank media was used as a control. Plates undergoing TGF- β stimulation were incubated for 1 hr in the lower serum media before compound addition. TGF- β stimulation was carried out at 20 ng/ml. Time point images were collected between 0 hr and 44 hr until scratches were completely healed. Images were analysed using MRI wound healing macro plugin for Image J (Baecker, 2011; Schneider *et al.*, 2012; Schindelin *et al.*, 2012). Graphs and statistical analysis was carried out using GraphPad Prism 7.

2.4. NMR-based procedures

2.4.1. NMR sample preparation

NMR protein samples were dialysed into NMR buffer (20 mM Na₂HPO₄, 50 mM NaCl pH 6.8) and spin concentrated to 0.5 – 1 mL using a Vivaspin 20 spin concentrator (Sartorius) at 4300 x g 4 °C (Heraeus Megafuge 1.0R). The sample was then prepared as described in Table 2.4.1, filtered using a 0.22 µm spin filter (Corning) and transferred to an NMR tube (5mm Wilmad 535PP).

Component	Volume
Protein sample	500 µL
1.5 % sodium azide	11.7 µL
DSS	11.7 µL
D ₂ O	58.1 µL

Table 2.4.1 - NMR sample preparation components and volumes.

DSS - 4,4-dimethyl-4-silapentane-1-sulfonic acid.

2.4.2. NMR spectroscopy experiments

All experimental data was collected using either a Bruker Avance III 800 MHz spectrometer or a Bruker Avance I 500 MHz spectrometer equipped with triple resonance probes at 298 K. For the GB1-WW1 backbone assignment, experiments included: [¹H-¹⁵N]-HSQC, HNCACB and CBCA(CO)NH as described in Table 2.4.2. The backbone resonance assignments for Smad-7 peptide bound-GB1-WW2 were obtained from Danielle Bourcier. For GB1-WW4, all experiments collected are described in Table 2.4.2 with a sample concentration of 2 mM. Other experiments and backbone assignments used for structural analysis of GB1-WW4 were collected by Lloyd Wahl (2016). The aro-NOESY (¹H-¹H 2D plane NOESY-HSQC at 120 ppm for ¹³C) had a mixing time of 50 ms, whereas the ¹³C-NOESY-HSQC and ¹⁵N-NOESY-HSQC acquired by Lloyd Wahl (2016) had mixing times of 100 ms. Fluorine spectra were also run on two titration samples to help understand the results. These were carried out in a Bruker Avance III 500 MHz spectrometer with a broadband probe.

Experiment	Scans	Complex points		Spectral widths		
		¹⁵ N	¹³ C	¹ H	¹⁵ N	¹³ C
[¹ H- ¹⁵ N]-HSQC	8	128	-	12019	2594.7	-
HNCACB	32	32	64	12019	2594.7	15105.7
CBCA(CO)NH	32	32	53	12019	2594.7	15105.7

Table 2.4.2 - Acquisition parameters for GB1-WW1 NMR experiments.

All experiments were carried out in the 800 MHz spectrometer. All experiments used 1024 complex points for ¹H.

Experiment	Scans	Complex points		Spectral widths		
		¹⁵ N	¹³ C	¹ H	¹⁵ N	¹³ C
[¹ H- ¹⁵ N]-HSQC ^b	8	128	-	7507.5	1620.4	-
HCCH-TOCSY ^b	8	120(¹ H)	32	7507.5	6995.6(¹ H)	9423.8
CCH-TOCSY ^b	8	120(¹³ C)	32	7507.5	9423.8(¹³ C)	9423.8
Aro-NOESY ^a (¹ H- ¹ H 2D plane NOESY-HSQC @ 120 ppm for ¹³ C)	16	256(¹ H)	0	12019	10395(¹ H)	15105.7
HbCbCgCdHd ^a	256	-	48	12820.5	-	6038.6

Table 2.4.3 - Acquisition parameters for GB1-WW4 NMR experiments.

^a indicates experiments carried out in the 800 MHz spectrometer and ^b in the 500 MHz spectrometer. All experiments used 1024 complex points for ¹H.

2.4.3. NMR titrations

All experimental data was collected from the same spectrometers as described in section 2.4.2. NMR titrations were run using ¹⁵N-labelled GB1-WW1 and GB1-WW1-2 and synthetic peptides based on Smad proteins 2, 3 and 7. The synthetic peptides were purchased from Proteogenix and Peptide Synthetics. Synthetic peptide sequences are described in Table 2.4.4. GB1-tagged protein samples were prepared as described in Table 2.4.1. Titrations were carried out using concentrated peptide in μ L amounts as described in Tables 2.4.5 – 2.4.10.

Peptide	Amino acid sequence
(SUMO)-Smad 2	IEPQSNYIPETPPPGYISEDGE
(SUMO)-Smad 3	IEPQSNIPETPPPGYLSEDGE
Smad 2	IPETPPPGYISEDGE
Smad 3	IPETPPPGYLSEDGE
Smad 7	ELESPPPPYSRYPMD

Table 2.4.4 - Residue sequences for the recombinant peptides and synthetic peptides used in expressed-peptide generation and NMR titration analyses, respectively.

Molar Ratio	[Smad 2 peptide] (mM)	Volume Smad 2 peptide added (μ L stock – 100 mM)
1:0	0	0
1:0.12	0.07	0.4
1:0.23	0.14	0.4
1:0.46	0.28	0.8
1:0.92	0.55	1.6

Table 2.4.5 - Conditions used during GB1-WW1 titration with synthetic Smad 2 peptide.

A concentration of 0.6 mM 15 N-GB1-WW1 was used in this titration.

Molar Ratio	[Smad 3 peptide] (mM)	Volume Smad 3 peptide added (μ L stock – 100 mM)
1:0	0	0
1:0.13	0.08	0.4
1:0.26	0.16	0.4
1:0.5	0.32	0.8
1:1	0.64	1.6
1:2.1	1.24	3.2
1:4.1	2.43	6.3
1:6	3.59	6.3
1:7.9	4.73	6.3
1:9.7	5.84	6.3

Table 2.4.6 - Peptide additions and concentrations throughout the GB1-WW1 titration using synthetic Smad 3 peptide.

A concentration of 0.6 mM 15 N-GB1-WW1 was used in this titration.

Molar Ratio	[Smad 7 peptide] (mM)	Volume Smad 7 peptide added (μ L stock – 10/100 mM)
1:0	0	0
1:0.08	0.005	0.3
1:0.17	0.010	0.3
1:0.43	0.026	0.9
1:0.87	0.052	1.5
1:1.7	0.10	2.9
1:3.4	0.20	<i>0.59</i>
1:5.1	0.31	<i>0.59</i>
1:6.8	0.41	<i>0.59</i>
1:8.5	0.51	<i>0.59</i>

Table 2.4.7 - Conditions used during the GB1-WW1 titration with synthetic Smad 7.

Italics refer to additions made using 100 mM rather than 10 mM. Titration carried out with Dr Lloyd Wahl. A concentration of 0.06 mM 15 N-GB1-WW1 was used in this titration.

Molar Ratio	[Smad 2 peptide] (mM)	Volume Smad 2 peptide added (μ L stock – 100 mM)
1:0	0	0
1:0.125	0.09	0.5
1:0.25	0.18	0.5
1:0.5	0.36	1.0
1:1	0.71	2.0
1:2	1.43	3.95
1:4	2.86	7.85
1:6	4.29	7.85
1:8	5.72	7.85
1:10	7.15	7.85

Table 2.4.8 - GB1-WW1-2 titration conditions using synthetic Smad 2 peptide.

A concentration of 0.7 mM 15 N-GB1-WW1-2 was used in this titration.

Molar Ratio	[Smad 3 peptide] (mM)	Volume Smad 3 peptide added (μ L stock – 100 mM)
1:0	0	0
1:0.2	0.07	0.4
1:0.5	0.14	0.4
1:0.9	0.28	0.8
1:1.6	0.48	1.2
1:2.4	0.72	1.4
1:4.9	1.47	4.3
1:7.4	2.21	4.3
1:9.8	2.95	4.3

Table 2.4.9 - Concentrations used, and peptide additions during GB1-WW1-2 titration with synthetic Smad 3 peptide.

A concentration of 0.3 mM 15 N-GB1-WW1-2 was used in this titration.

Molar Ratio	[Smad 7 peptide] (mM)	Volume Smad 7 peptide added (μ L stock – 100 mM)
1:0	0	0
1:0.13	0.02	0.1
1:0.27	0.04	0.1
1:0.5	0.08	0.2
1:1	0.15	0.4
1:2.26	0.34	0.9
1:4.6	0.69	1.75
1:6.9	1.04	1.75
1:9.2	1.39	1.75
1:11.6	1.74	1.75

Table 2.4.10 - Smad 7 peptide additions and concentrations in GB1-WW1-2 titration.

A concentration of 0.15 mM 15 N-GB1-WW1-2 was used in this titration.

2.4.4. Processing and analysis of spectra

Spectra were processed using NMRPipe software (Delaglio *et al.*, 1995) in which phasing, zero filling, linear prediction and baseline corrections where necessary were applied to spectra. Analysis of spectra was carried out in CCPN analysis (Vranken *et al.*, 2005).

2.4.5. Structural calculations

Dihedral restraints were calculated using DANGLE (Cheung *et al.*, 2010) set to reject predictions with over 2 islands. Predictions with 2 regions were either accepted or rejected based on the Φ and Ψ boundary limits, generating an initial 192 restraints. Distance restraints were calculated within CCPN analysis using the unassigned peaks in the ^{13}C -NOESY-HSQC, ^{15}N -NOESY-HSQC and a 2D ^1H - ^1H plane of a ^{13}C -HSQC-NOESY, with ^{13}C centred at 120 ppm (aro-NOESY). Shift match restraints used the average peak volume as representative of 3.2 Å, which was then used to generate a list of distance restraints. These restraints lists were split into ambiguous and unambiguous. The ccpn project was then submitted to the CCPN NMR grid (Fogh *et al.*, 2005; Nilges *et al.*, 2008; Rieping *et al.*, 2007; Vranken *et al.*, 2005) to run ARIA-based structural calculations. A total of 8 iterations were carried out with 20 initial models that were then refined in explicit water to 10 refined models. The updated project which was analysed in CCPN Analysis which contained a structural ensemble, a list of peak assignments and new restraints lists with violations. Violations with a mean violation of 0.2 or higher were analysed within the project and their source (misassignments, for example) corrected where possible, before another set of restraints were generated using the new assignments and the project submitted again for structural calculations. 27 runs were carried out in this manner. When no more corrections could be made using this method, further calculations were carried out using the WeNMR ARIA 1.2 HJ webportal (Linge *et al.*, 2003; Linge *et al.*, 2003; Linge *et al.*, 2001; Linge and Nilges, 1999; Nilges and O'Donoghue, 1998; Nilges, 1997; enmr.chemie.uni-frankfurt.de/portal/aria). This enabled the calculation of more structures per iteration and manual correction of restraint limits when restraints were at acceptable distances. Sequence, unambiguous/ambiguous/dihedral restraints, peak lists and chemical shift data were exported from the CCPN Analysis project after the 27th run, using CNS/NMRView formatting. The parameters and settings used in the WeNMR webportal are described in Table 2.4.11. The structural ensemble was visualised using Chimera (Pettersen *et al.*, 2004). New assignments and violations were checked against the original spectral data and corrected where possible. A further 49 runs were carried out this way, until no further violations with an average error above 0.3 Å were present.

Parameter	Setting
Number of structures per iteration	50
Number of structures in last cycle	100
Number of refined structures	20
Seed for the calculation	6789 (default)
¹³ C-NOESY/ ¹⁵ N-NOESY	
Proton 1	Tolerance 0.02
Hetero 1	Tolerance 0.50
Proton 2	Tolerance 0.02
NOESY mixing time	100 ms
Rotation correlation time	6 ns
Spectrometer frequency	800 MHz (¹³ C)/500 MHz (¹⁵ N)
Spin diffusion	Corrected
Peak assignments	Use all peaks and keep assigned peaks
Aro-NOESY	
Proton 1	Tolerance 0.02
Proton 2	Tolerance 0.02
NOESY mixing time	50 ms
Rotation correlation time	6 ns
Spectrometer frequency	800 MHz
Spin diffusion	Corrected
Peak assignments	Use only assigned peaks

Table 2.4.11 - Parameter settings used for the WeNMR ARIA 1.2 HJ webportal structural calculations.

2.4.6. Structural validation

Structural validation was performed using PSVS 1.5 (Bhattacharya *et al.*, 2007), Molprobit (Duke, US; Lovell *et al.*, 2003) and wwPDB (www.wwpdb.org; validate-rcsb-3.wwpbd.org, part of the Protein Data Bank; Berman *et al.*, 2003; Berman *et al.*, 2007).

3. Structural Characterisation of WW4

3.1. Introduction

Drug discovery utilises many different approaches with the aim of improving the specificity and efficacy of potential therapeutics. One such method involves knowledge of the target structure to design therapies based on protein folding and side chain positioning. This chapter looks at the structural solution of the fourth WW domain of the protein WWP2 using solution state NMR and how this can be used for peptide design.

NMR spectroscopy uses the relationship between particles and their ability to absorb and emit electromagnetic energy to manipulate an inherent property of atomic nuclei referred to as spin. Spin is a quantum property of subatomic particles that have an angular momentum, but this does not mean a rotational movement as the term may suggest (Keeler, 2010). A vector model is commonly used to describe this property, with both a direction and magnitude. Quantum spin is expressed in integer terms where both the neutron and proton each have a $\frac{1}{2}$ spin. Neutrons and protons combine to give the overall quantum spin value. Protein NMR utilises the half-integer spin properties of nuclei as these spectra are the easiest to interpret, with the simplest nuclei being ^1H with a total quantum spin of $\frac{1}{2}$, although deuterium is also used which has an integer spin of 1. Other commonly used isotopes are ^{13}C and ^{15}N , also with total quantum spin of $\frac{1}{2}$. However, as these only exist naturally to a very low abundance, in protein NMR incorporation into the protein is required and involves recombinant expression of the protein or selective labelling which can be expensive and time-consuming.

Within spin $\frac{1}{2}$ nuclei, there are two spin states, described by their position either with (up) or against (down) the magnetic field. In the absence of a magnetic field, the spin states across a sample randomly orientate between up and down positions resulting in degenerate orientations that have the same energy level. However, in the presence of a bulk magnetic field (B_0), the spin state energy levels change leading to a split into occupying either a lower energy state (α which aligns with the magnetic field) or a higher energy state (β which aligns against the magnetic field). The average of these states at equilibrium under B_0 create a net magnetisation vector along the z-axis of the spin vector model (Keeler, 2010) due to a tendency for more spins to exist at ground state. Additionally, after a set period of time, B_0 induces the magnetisation of the nuclei to precess about the z-axis at a fixed angle known as the angle of precession. Larmor frequency is the term used to describe the relationship between the strength of the external magnetic field and the gyromagnetic ratio of the nuclei of interest, and in turn

the frequency of precession of the nuclei magnetisation. This Larmor precession sets the basis for the rotational frame and vector model transitions that classically describe the nuclei during an NMR experiment.

The NMR signal that is collected during an experiment in its basic form develops via 3 stages (Keeler, 2010). Firstly, an initial period T_1 in the presence of the bulk magnetic field allows the sample spins to reach equilibrium; this is shortly followed by the application of a radiofrequency (RF) pulse at the frequency of the Larmor precession along the x or y axis, inducing a change in the magnetisation vector position away from the z-axis (the simplest being a change of 90°); and lastly the bulk magnetisation is measured from its precession along the x-y plane to its return to equilibrium. The signal collected from the change in magnetisation via the spectrometer coil is called the free-induction decay (FID). However, as there is usually more than one spin type that is detectable within a protein sample, each with different rates of Larmor precession, separate channels are used to individually excite specific nuclei at the same time. By using the proton for indirect detection of other nuclei, this utilises the higher gyromagnetic ratio of ^1H resulting in increased sensitivity and better spectra can be obtained. After the collection of many FID signals to increase the signal-to-noise ratio, the data is then converted into a form which relates to the frequencies of the nuclei. This is known as Fourier transformation, and presents the data as frequency peaks for easier analysis and interpretation.

As previously mentioned, the Larmor frequency of nuclei is dependent on the strength of the external magnetic field resulting in resonances having different frequencies under different magnetic field strengths. In order to compare chemical shift frequencies between experiments carried out using different strengths of magnet, Fourier transformed data is commonly described in parts per million, instead of frequency. This normalises the data so that the same atom that presents a peak at 500 Hz in a spectrum collected using a 500 MHz spectrometer, but has a frequency peak of 800 Hz in a spectrum acquired using an 800 MHz spectrometer, will result in both peaks residing at 1 ppm. Each atom will exist in slightly different chemical environments despite being of the same element, so the ppm value differs slightly for each atom and is referred to as the chemical shift. The chemical shift therefore corresponds to particular types of atoms in the surrounding chemical space of the sample, thus the type of protein NMR experiment can determine which chemical shifts are observed.

Different pulse sequences are used to determine distinctive aspects of the protein sample to build up an atomic picture of the protein. For assignment and structural studies, 2D and 3D experiments are used. The sequence of a standard 2D pulse sequence (outlined in Figure 3.1.1A) involves an initial preparation step in which coherence is reached by the equilibrium magnetisation (Keeler, 2010). This step can be either simple like a 90° pulse or more complex such as combinations of pulses used to transfer magnetisation to another nucleus. The coherence then processes and evolves during the evolution period, t_1 . This period can last for any time frame as it is determined by the number of increments in the experimental design and is important for labelling the magnetisation with the chemical shift of the indirect nucleus. After this period, the coherence is converted to an observable signal during the mixing period, which is also the step that determines the type of information collected from the experiment. The data is then recorded during t_2 , the detection period. So, during a 2D experiment the pulse sequence is initiated for a set number of increments, each with an increasing sampling period for t_1 (described in Figure 3.1.1B). This collected data is then Fourier transformed in both the t_1 and t_2 into the well-known 2D frequency spectra that can then be easily compared and analysed. The same principles apply to triple resonance experiments; however, these tend to be more complex on account of multiple types of through-bond connections transferring magnetisation to the nuclei of interest. The data collected from such experiments enables the sequential assignment of residues based on chemical shift values characteristic of the residue type and connecting residues. This is especially valuable if the spectra lack residue-specific peak positions. The combination of several triple resonance experiments linking the backbone with the side-chains, as well as identifying atoms in close proximity in space, builds a complex network of bonds and restraints into a 3D structure of the protein being observed.

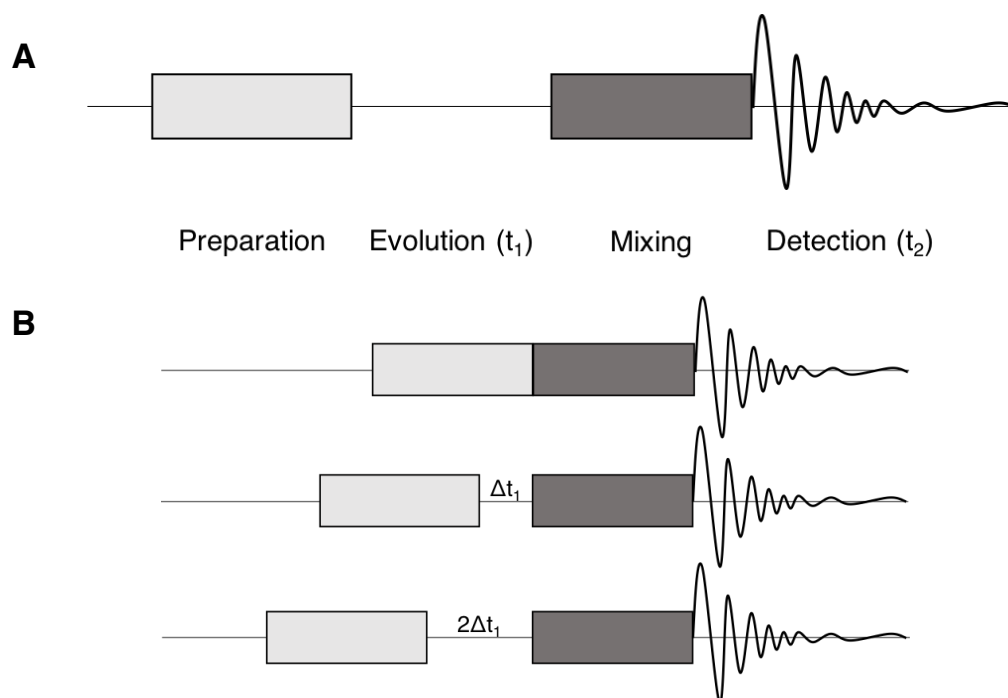


Figure 3.1.1 - Schematic representations of a general 2D NMR experiment.

A) Diagram representing the general composition of a 2D NMR experiment. B) Diagram showing the principle of recording FIDs with increasing increments of a 2D NMR experiment. Light grey box refers to the preparation period; dark grey box represents the mixing period. Adapted from Keeler (2010).

The heteronuclear single-quantum correlation (HSQC) experiment is one of the principal 2D NMR experiments for the analysis of proteins. Heteronuclei, for example ^1H and ^{15}N or ^1H and ^{13}C , which are directly bound generate cross peaks from the transfer of magnetisation from one nucleus to the other via J-coupling (Keeler, 2010). In simplified terms, the pulse sequence involves magnetisation initially being transferred from the first spin (^1H) to the second (^{13}C or ^{15}N) using a series of pulses called INEPT (insensitive nuclei enhanced by polarisation transfer). The second spin then evolves for t_1 before magnetisation is transferred back to the first spin which is subsequently observed. By directly detecting the proton, the highest sensitivity can be achieved as the nitrogen/carbon signal is much weaker. In combination with isotopic labelling of the other nucleus (^{13}C or ^{15}N), this provides clean peak separation across both dimensions, where each dimension corresponds to the different nuclei. This produces single peaks for each amide moiety of the protein sequence, the exception being proline as it lacks the amide. Additional peaks for tryptophan and arginine $\text{N}\epsilon\text{H}\epsilon$ can also be observed although the latter tends to appear as aliased peaks due to being outside the spectral region. However, the NH_2 side-chains of asparagine and glutamine

can be observed as two peaks at different ^1H shift values but the same ^{15}N chemical shift position. Taken together with triple resonance data sets, peaks can be assigned to individual residue atoms of the protein sequence. In addition to identifying the individual amide groups in the protein, other aspects of the sample can also be determined, such as whether the protein concentration is high enough to produce a good signal-to-noise ratio or identify regions of the protein that may be unfolded. Furthermore, the influence of other factors on the protein such as temperature, pH and the presence of ligand can be observed. It is due to these reasons that a HSQC experiment is most commonly one of the first spectra to be acquired for a protein sample and can determine whether further characterisation by protein NMR is worthwhile.

In the case of WWP2, substrate selection is carried out by four WW domains. As mentioned earlier, WW domains are composed of 30-40 amino acids with two tryptophan residues and a highly conserved proline residue (Kasanov *et al.*, 2001; Panwalker *et al.*, 2017). These domains interact with a PPxY motif via an anti-parallel β -sheet fold consisting of three strands. The positioning of aromatic residues on the surface of the β -sheet has been shown to create grooves known as XP grooves, which facilitate the binding of the substrate. Furthermore, key residues in other parts of the structure have been found to affect the affinities of the WW domain towards particular substrates (Kasanov *et al.*, 2001). Interestingly, the fourth WW domain within WWP2 lacks the 2nd tryptophan residue which is replaced by phenylalanine but seems to retain other highly conserved residues, described in chapter 1, which are required for function. Understanding the structure of particular WW domains and the orientations of key residues can provide vital information on the functionality and the selection of substrates.

Knowing the structure of the protein is useful in many regards. In drug design, it is particularly useful in showing the arrangement of side chains that are involved in interactions or form the binding pocket, or in the case of bound structures, highlights key areas of conformational change induced by binding that may not necessarily be within the residues directly involved in binding. Therefore, both unbound and bound structures are information rich and can be used for more intelligent and specific design of compounds or peptides. In order to enable inhibitor design using a structure-based approach, the aim of this chapter was to solve the WW4 domain conformation using solution state NMR. Previous work by Dr Lloyd Wahl (2016) using ^{13}C - ^{15}N -GB1-WW4

presented a partially refined structure for WW4 of WWP2, however further assignment of atoms was required in order to produce a fully refined ensemble for WW4. Building upon the data from Dr Lloyd Wahl, the addition of further experiments in this thesis, to increase the unambiguous and more complete assignment of WW4 resonances, allowed for the automated assignment of NOESY-based restraints, and iterative structural calculations using the CCPN Grid (Boucher *et al.*, 2005; Nilges *et al.*, 2008; Rieping *et al.*, 2007; Vranken *et al.*, 2005) and WeNMR ARIA 1.2 HJ webportal (<http://enmr.chemie.uni-frankfurt.de/portal/index.html>).

3.2. Results

The use of the solubility tag GB1 enabled the expression and purification of WW4 in isotopically labelled media for use in NMR experiments. The tag was not cleaved before experimentation as previous work showed removal resulted in protein precipitation (Lloyd Wahl, 2016; unpublished observations).

3.2.1. HSQC analysis for sample comparison of double labelled WW4

In order to acquire additional triple resonance experiments for use alongside those collected by Dr Lloyd Wahl (2016), a double-labelled sample of GB1-WW4 at 2 mM in sodium phosphate buffer was produced using bacterial expression and protein purification methods described in sections 2.2.2 and 2.2.4. Initial HSQC analysis of the sample looked as expected when compared to the previous GB1-WW4 HSQC spectrum (Wahl, 2016) (Figure 3.2.1). Hence further acquisition of triple resonance experiments was carried out using the 800 MHz spectrometer.

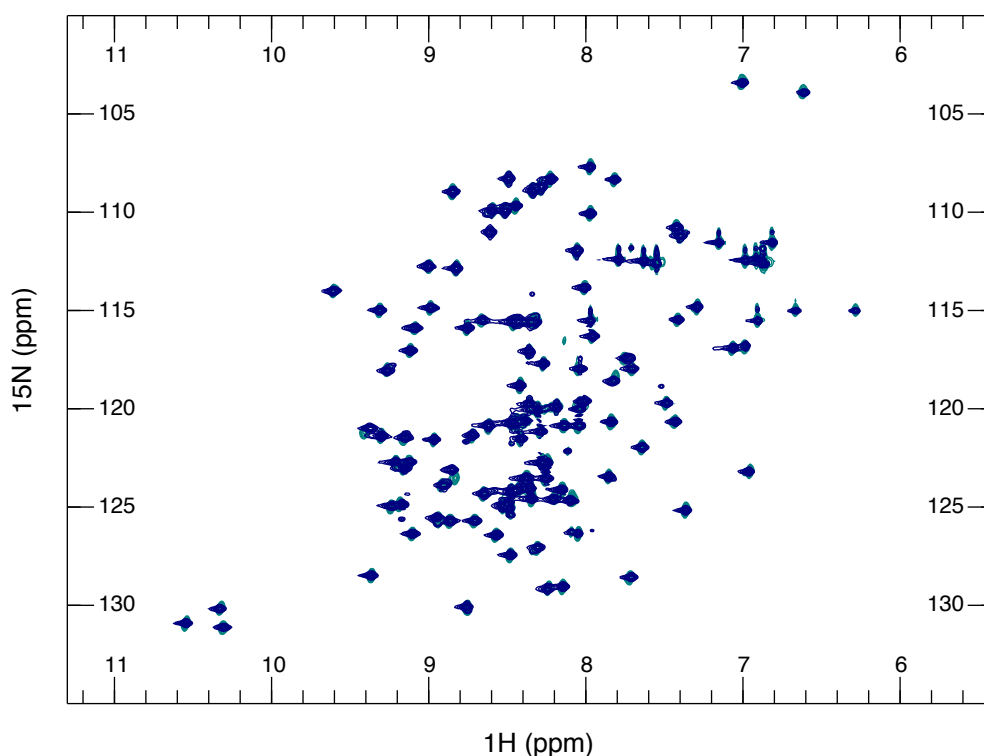


Figure 3.2.1 - Comparison of double-labelled ^{13}C - ^{15}N -GB1-WW4 samples by HSQC.

Overlaid HSQCs of Lloyd Wahl's WW4 sample (teal) used for previous data collection, and the HSQC of the sample used in this research (dark blue) for additional experiment acquisition. ppm – parts per million.

3.2.2. Assignment of GB1-WW4 side chains

Previous work towards the structural solution of WW4 (Wahl, 2016) was built upon by obtaining further triple resonance experiments for side chain assignment. Spectral data was transformed and processed using NMRPipe and NMRDraw (Delaglio *et al.*, 1995). The HCCH-TOCSY and CCH-TOCSY were obtained and assigned for aliphatic side chain resonances, and the HBCBCGCDHD and a 2D ^1H - ^1H plane of a ^{13}C -NOESY-HSQC with the ^{13}C centred at 120 ppm (aro-NOESY) were acquired and assigned for aromatic side chain resonances, using CCPN Analysis (Vranken *et al.*, 2005).

3.2.3. ARIA: automated peak assignment, structural calculations and water refinement

For structural calculations, assignment of nuclear Overhauser effect (NOESY) spectra peaks determines which residue atoms were within 5-6 Å of each other. It is possible

to manually assign each peak from other resonance values that have already been assigned in other spectra. However, as this stage in data analysis is the most time-consuming part due to numerous possible assignments, the presence of potential noise peaks/solvent peaks and proton overlaps, there are now automated methods of NOESY peak assignment which can overcome such obstacles.

After manual peak picking in the ^{13}C -NOESY-HSQC, ^{15}N -NOESY-HSQC and aro-NOESY spectra to facilitate automated assignment at a later stage, the next step was to generate distance and dihedral restraints which provided information about the spatial arrangement of atoms. These restraints were generated using DANGLE (Dihedral ANGles from Global Likelihood Estimates; Cheung *et al.*, 2010) and 'Structure: Make Distance restraints' within CCPN Analysis. DANGLE uses a combination of the molecule sequence, the chemical shift data present in the CCPN project, and comparison to known protein shifts and structures to predict the probability of given backbone orientations in space. A total of 192 Φ and Ψ backbone restraints were calculated using a maximum setting of 2 predicted islands, and all other settings at default. Any residues with over 2 islands predicted in Ramachandran space were automatically removed, and those outside of the Φ and Ψ upper and lower limits were also discounted resulting in dihedral angles for 98 of 109 residues. As the outliers were mainly at the N-terminal regions and flexible linker between amino acids 62 and 73, this was deemed adequate for further progression to structural calculations. As for distance restraints, 123 unambiguous and 2420 ambiguous distance restraints were initially generated within CCPN Analysis. The software utilised average peak volume as an intensity reference for a distance of 3.2 Å (Skinner *et al.*, 2015), and then derived individual NOE distances from this value based on the specific intensity values for each peak. After the restraints were created, the lists were merged and split according to whether they were unambiguous or ambiguous ready for use in structural calculations.

The project was then submitted to the CCPN grid for the next stage of structure generation. The CCPN NMR grid (Fogh *et al.*, 2005; Nilges *et al.*, 2008; Rieping *et al.*, 2007; Vranken *et al.*, 2005) was used for automated assignment and structural calculations up to a total of 27 runs. This portal used a combination of the CCPN data model, ARIA 2.1 and CONDOR and enabled an alternative way of using ARIA-based iterative calculations rather than via a local source due to previous problems with local versions of ARIA. Within each run, there were 8 iterations with 20 models, and a final ensemble of 10 models with the lowest energies were produced after water refinement.

For example, the first ensemble from run 1 for the GB1 and WW4 regions shown in Figure 3.2.2A and B, respectively. All assignment, restraint and structural data was retrieved as a CCPN project and analysed in Analysis. A total of 2652 peaks were present in the final iteration of the first CCPN grid run, compared with 3091 in the original input lists. This suggested the presence of erroneous peaks - peaks that could not be assigned by ARIA, potentially due to ambiguity, overlap or lack of any assignments for the given atoms in the project. The rejected restraints (i.e. violations), which were not compatible with the structures calculated, were checked and corrected from the most severe (red) to the least (yellow), where possible. After 27 runs using the CCPN grid, violations were still present, in which the majority appeared to be within the absolute distance limits with correct assignments based on the spectral connections analysed. At this stage, it was decided that changing software to one which would generate a greater number of models per iteration and allow manual correction of restraint distance limits, within acceptable bounds, may be more beneficial for analysing and correcting the remaining errors, enabling improvement towards a better-quality structure.

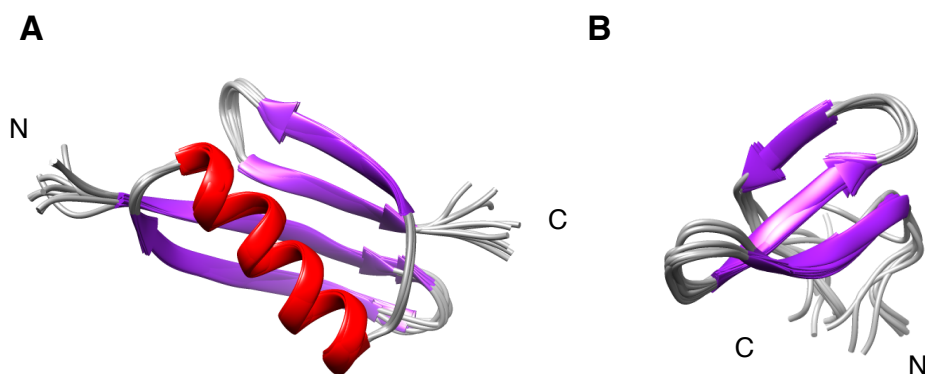


Figure 3.2.2 - Initial partially refined GB1 and WW4 ensemble alignment from CCPN grid run 1.

A) GB1 ensemble (1 – 61) consisting of 10 models with the helix in red, β -strands in purple and random coils in grey. B) WW4 ensemble (73 – 106) of 10 refined models with the β -strands in purple and random coils in grey. Residues 62 – 72 are not shown as the region was disordered and prevented sufficient alignment of the independent domains for analysis. Images were created using Chimera (Pettersen et al., 2004). N and C refer to the amino and carboxyl termini.

Sequence, chemical shifts, ARIA-based peak lists and restraints data were exported from CCPN Analysis from the last CCPN Grid result project to facilitate the use of the WeNMR ARIA 1.2 HJ webportal for further assignment, structural calculations and

refinement. 50 structures were generated per iteration with 100 structures in the last cycle. A final set of 20 models were generated after refinement in explicit water. A total of 49 runs were carried out using this portal. The final iteration of the final run used a total of 1803 restraints, with the distribution of 873 intra-residue, 400 sequential, 118 medium range and 412 long range. Minimal restraints were present for the linker region between GB1 and WW4, while no restraints spanned GB1 to WW4, highlighting each domain as independently folding within the sample. The final refined ensemble can be seen in Figure 3.2.3A, where the GB1 domain has been aligned. The same overall secondary structures for the individual GB1 (Figure 3.2.3B) and WW4 (Figure 3.2.3D) domains are mirrored in this ensemble when compared to the initial structural models shown in Figure 3.2.2 A and B respectively, showing that this folding pattern is derived from the NOEs and restraints derived from the experimental data. The backbone atom root mean standard deviation (RMSD) for this final ensemble over residues 1-109 is 8.9 Å, with a heavy atom RMSD at 9.1 Å. This value is very high for a final structure, however due to the highly flexible, ill-defined linker region between the two domains with very few restraints, it is not surprising as this linker results in many different spatial arrangements of WW4. When restricted to just the well-defined core regions of GB1 and WW4 consisting of residues 3-59 and 75-101 respectively (as calculated by wwPDB validation service, discussed later) the RMSD values for the backbone atoms was much lower at 0.4 Å for both domain regions and the heavy atoms RMSD reduced to 0.8 Å and 0.9 Å for GB1 3-59 and WW4 75-101 respectively (calculated by PSVS). This suggests that the variation between the models of these core regions is low and supports this ensemble being a good quality model when considering each of the individual core regions. Furthermore, analysis of the GB1 tag compared to the published structure of PDB code 3GB1 (Juszewski *et al.*, 1999; Berman *et al.*, 2000), shown in Figure 3.2.4, showed only minimal differences between the tag structures. Overall, each domain presented with a medoid model which refers to the model that reflects the most similarity to the other structures in the ensemble. These are shown in Figure 3.2.3C for GB1 (model 16) and Figure 3.2.3E for WW4 (model 7).

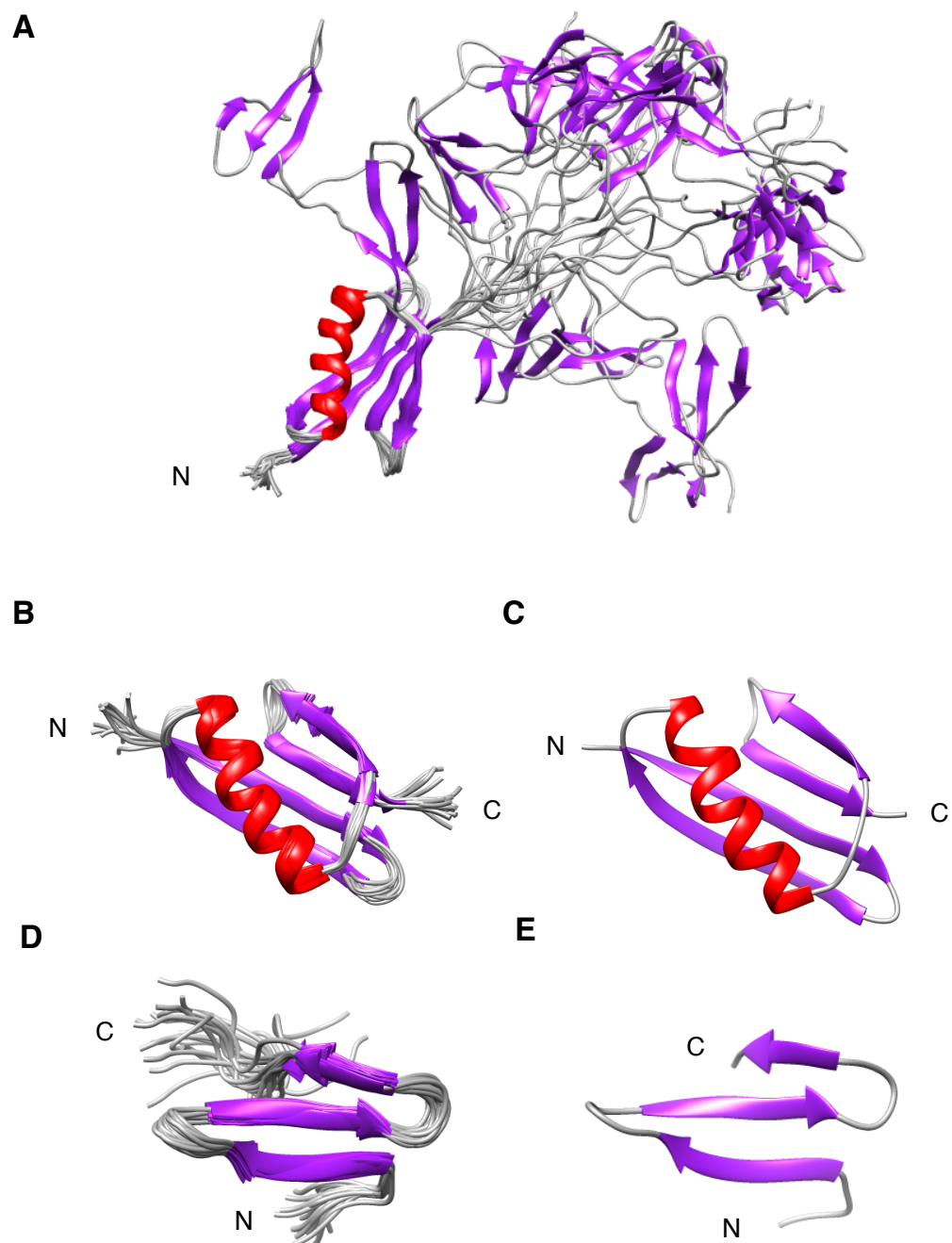


Figure 3.2.3 - Final refined ensemble for GB1-WW4 from the WeNMR ARIA 1.2 HJ webportal calculations.

A) 20 refined models for GB1-WW4 amino acids 1-109 with GB1 region aligned. B) GB1 ensemble ranging from amino acid 1 to 61 consisting of 20 models. C) Most representative model (number 16) of GB1 as defined by wwPDB validation analysis, from amino acid 3 to 59. D) 20 refined models for WW4 from residue 73 to 109. E) Most representative model for WW4 (model number 7) from amino acid 75 to 101. Images were generated using Chimera (Pettersen et al., 2004). N and C refer to the amino and carboxyl termini.

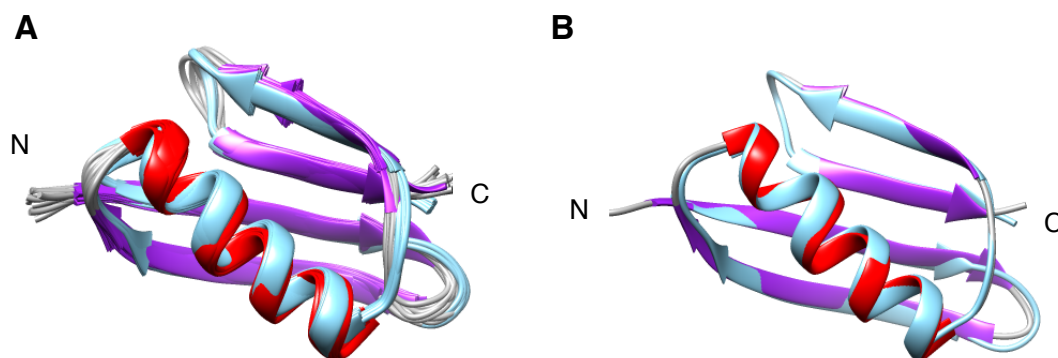


Figure 3.2.4 - Ensemble comparison of GB1 in GB1-WW4 with published 3GB1.

A) 32 3GB1 models (light blue) aligned with 20 refined GB1 models (red – helix; purple - β -strands). B) Most representative model of GB1 in GB1-WW4 ensemble aligned with the first model of PDB 3GB1. Alignments and images generated using Chimera (Pettersen *et al.*, 2004). N and C refer to the amino and carboxyl termini.

3.2.4. Structure Validation

In order to truly assess the quality of the final ensemble generated by structural calculations and refinement discussed in the previous section, structural validation software PSVS 1.5 (Bhattacharya *et al.*, 2007), MolProbity (Duke, US; Lovell *et al.*, 2003) and wwPDB (Berman *et al.*, 2003; Berman *et al.*, 2007) were employed.

PSVS utilises evaluation tools such as PROCHECK (Laskowski *et al.*, 1993), MolProbity (Lovell *et al.*, 2003), Verify3D (Luthy *et al.*, 1992) and Prosa II (Sippl, 1993) to look at how well the structure fits with already published X-ray structures, and the experimental data. The global quality scores give an indication of how well the ensemble performs based on certain criteria: Verify3d assesses the probability that the 3D structure reflects the predicted amino acid sequence-based spatial conformations; Prosa II looks at β -carbons close together spatially to derive the energy value of the interaction; Procheck analyses the dihedral angles ϕ (phi) and ψ (psi) in the backbone (Ramachandran plot analysis) as well as overall stereochemistry analysis to determine the quality of the ensemble; and lastly, MolProbity identifies any bad contacts which result in steric clashes or overlap of atoms, analyses how favourable the β -carbons spatial arrangements with regards to the backbone conformation and side-chain orientations as well as Ramachandran plot analysis. A summary of the PSVS output is shown in Table 3.2.2. As previously mentioned, the RMSD values for the backbone and heavy atoms when analysing all residues of the construct is high, but when only

considering the core regions, reduces to an acceptable value for each domain ensemble suggesting they are reliable representations of the structural architecture. Ramachandran plot analysis was carried out using both Procheck and MolProbity within PSVS. By plotting the ϕ and ψ angle values against each other, the regions of distinct secondary structures and the residues which adopt these particular ideal angles can be identified. Likewise, outliers, angles for residues which fall outside of the ideal values, can also be highlighted. For the core regions analysed by PSVS, as outlined in Table 3.2.2, no dihedral angles were disallowed, with the majority residing in favourable conformations, suggesting that the ensemble portrays structural angles in the ideal region. The webserver for MolProbity was also used to analyse all residues from 1-109 with respect to the dihedral angle acceptability. The resulting plots can be seen in Figure 3.2.5. 92.7 % were favourable, 6.7 % within the allowed boundaries, and 0.6 % disallowed. All the outlying residues resided either within the flexible linker or at the N-terminal region of the construct. This is potentially due to these regions being unable to produce NOEs because they are disordered and/or solvent exposed. While the values in Table 3.2.2 and Figure 3.2.5 do indicate a good level of quality, it does not completely eliminate the possibility of errors with bond lengths, angles, chirality and planarity within the structures. Overall, however, the ensemble performs well showing the structural model is of good quality.

Criteria	GB1-WW4 1-109	GB1 3-59	WW4 75-101
RMSD			
Backbone atoms	8.9 Å	0.4 Å	0.4 Å
Heavy atoms	9.1 Å	0.8 Å	0.9 Å
Bond angles	0.8°	0.9°	0.8°
Bond lengths	0.008 Å	0.008 Å	0.009 Å
Global quality scores	(raw/z-score ^a)	(raw/z-score ^a)	(raw/z-score ^a)
Verify3D	0.36 -1.61	0.43 -0.48	0.27 -3.05
Prosa II (-ve)	0.61 -0.17	0.80 0.62	0.30 -1.45
Procheck (phi/psi) ^b	-0.35 -1.06	-0.26 -0.71	-0.34 -1.02
Procheck (all) ^b	-0.42 -2.48	-0.38 -2.25	-0.35 -2.07
MolProbity clashscore	13.02 -0.71	16.43 -1.29	9.63 -0.13
Ramachandran plot -Procheck	(3-59,71-73,75-107)		
Most favoured	90.4 %	93.4 %	86.3 %
Allowed	9.6 %	6.6 %	13.7 %
Generously allowed	0.0 %	0.0 %	0.0 %
Disallowed	0.0 %	0.0%	0.0 %
Ramachandran plot -MolProbity	(3-59,71-73,75-107)		
Most favoured	98.5 %	99.2 %	97.2 %
Allowed	1.5 %	0.8 %	2.8 %
Disallowed	0.0 %	0.0 %	0.0 %

Table 3.2.2 - Summary of PSVS validation of GB1-WW4 and individual GB1 and WW4 core regions.

PSVS validation analysis of the final GB1-WW4 ensemble of 20 models. Regions analysed by PSVS are outlined either in the column heading or in brackets. ^a relative to 252 X-ray structures with < 500 residues and resolution of ≤ 1.8 Å, R-factor ≤ 0.25 and R-free ≤ 0.28 . Positive values indicate a better than average score. Adapted table from PSVS results report. RMSD – root mean standard deviation.

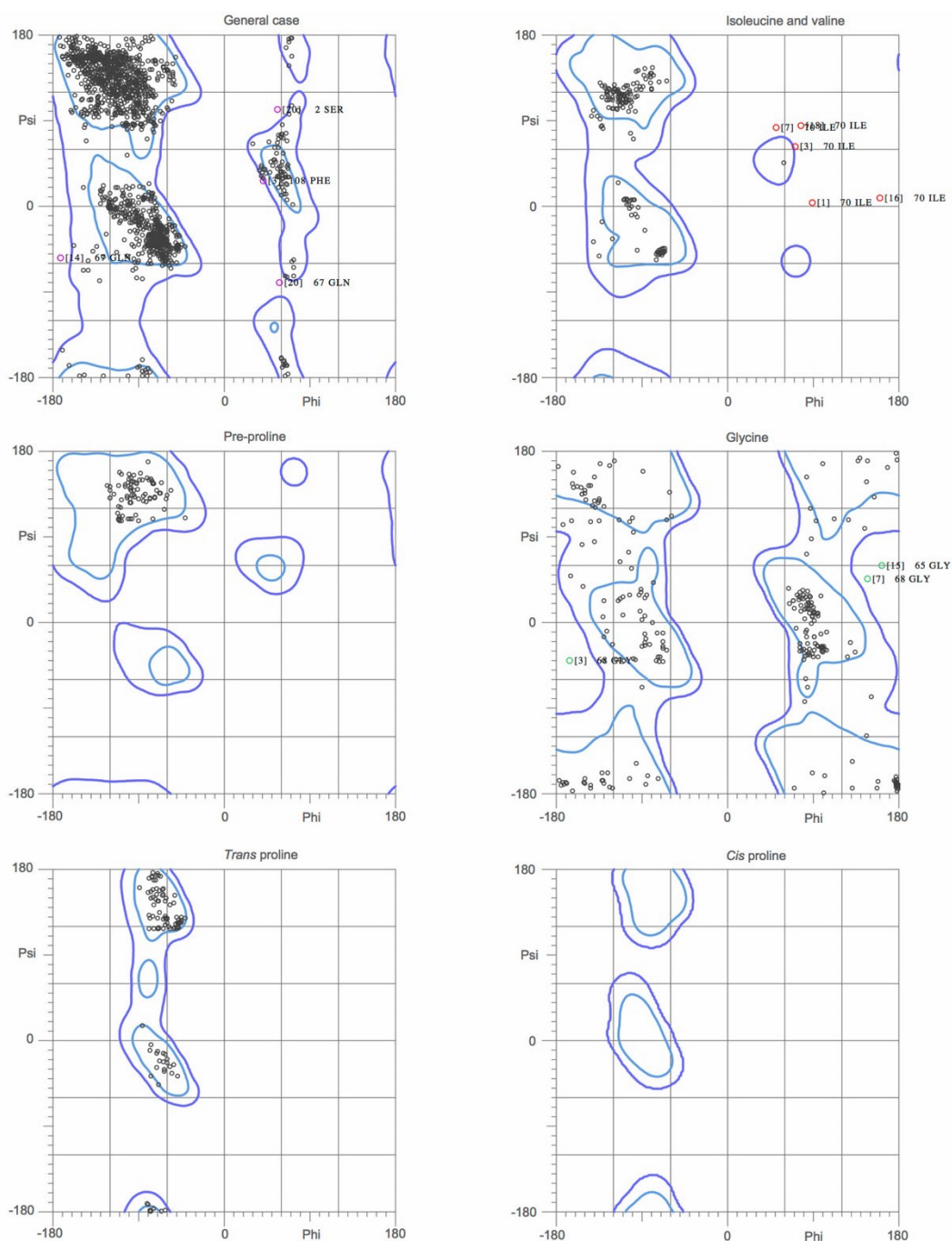


Figure 3.2.5 - MolProbity Ramachandran plots representing the analysis of 20 water refined models of GB1-WW4.

Structure quality analysis by Ramachandran plot for amino acids 1-109 across the final 20 water refined models of GB1-WW4. Light blue corresponds to favourable regions, whereas dark blue highlights allowed regions. Outliers shown as pink (general amino acids) /red (isoleucine and valine) /green (glycine) dots with model number and residue annotated. 92.7 % of all residues within the favoured regions, and 99.4 % within the allowed regions. Analysis carried out using MolProbity webserver (Duke, US; Lovell et al., 2003).

The final piece of validation software used for the final ensemble of 20 GB1-WW4 refined models was the One Dep wwPDB Validation service. The wwPDB is the place where final structural ensembles are deposited. However, these validation tests can also be run separately from deposition as a standalone validation service to determine the quality of the protein ensemble according to the deposition criteria at any time. The report presented an overall summary using a percentage slider scale for the ranking of clash score, Ramachandran outliers and side chain outliers (MolProbity; Chen *et al.*, 2010), in comparison with either all deposited structures (both crystal and NMR structures), or just NMR structures. GB1-WW4 scored well compared to all structures except for side-chain outliers that presented a slightly 'worse' score at 6.4% (Figure 3.2.6A). However, when compared to only NMR structures, GB1-WW4 ranks for all criteria were on the 'better' side of the scale. This suggests that the ensemble is well refined.

The wwPDB validation software only validates the well-defined regions of the protein PDB ensembles derived from NMR data that are submitted (Kirchner and Güntert, 2011; Montelione *et al.*, 2013). In this case, only residues 3-59 and 75-101 were included in the analysis parameters which mirrors the fact that the construct consists of 2 independent domains – the initial being the GB1 solubility tag and the latter being the domain of interest WW4. This is further supported by the random coil analysis carried out by the wwPDB (Figure 3.2.6B) showing 60-74 and the C-terminus as disordered. The validation of the GB1 domain supports the quality of the calculated ensemble from the NMR data, as another solution structure of GB1 has already been deposited to the PDB providing a good quality structure for comparison, as shown previously in Figure 3.2.4. Furthermore, as the GB1 structures align very well between the published and the GB1-WW4 ensemble, this highlights the absence of effect from the presence of GB1 on the overall structure of WW4 and vice versa other than improving the solubility of the construct. Analysis of the chain geometric quality based on outlier presence (red/orange/yellow) for individual amino acids, shown in Figure 3.2.6C and D, also indicates a good quality structural model. Investigation of the ensembles well-defined regions by the wwPDB is summarised in Table 3.2.3, and further reflects that the ensemble is well-refined and in good agreement with the other validation results. There were a few outliers regarding bond lengths, all within aromatic residue side-chains, probably from a missing assignment or loose/tight restraint. Chemical shift validation was also carried out by the wwPDB validation service on the

final ensemble, with Table 3.2.4 outlining the completeness of assignment for GB1-WW4. A total 1249 chemical shifts were analysed, with only 9 outliers identified. Analysis of the outliers suggest that the unusual shifts are most likely the result of nearby aromatic rings being close enough that the magnetic field generated by the delocalised electrons of the ring have influenced the frequency of the observed nuclei. Furthermore, comparison of the ppm values for the GB1-based outliers are consistent with those submitted to the BMRB for GB1 (PDB: 3GB1), supporting the conclusion that these are correct chemical shift values. As for the completeness of assignments, this assessed the number of assigned peaks in the entire structure and the well-defined core regions compared to the total number of assignable nuclei (Montelione *et al.*, 2013). Briefly, this includes backbone atoms C', N, HN and H α , side chain aliphatic nuclei from C β including carbon, hydrogen and nitrogen atoms, and all nuclei of the aromatic rings. Considering that the OH of tyrosine rings and CO groups were not assigned, and that the first few residues could not be seen, the overall percentage of assigned nuclei is good. Collectively, the wwPDB validation results showed that despite the ensemble having a few less-than-ideal aspects, the ensemble is well refined and is in good agreement with the other validation results.

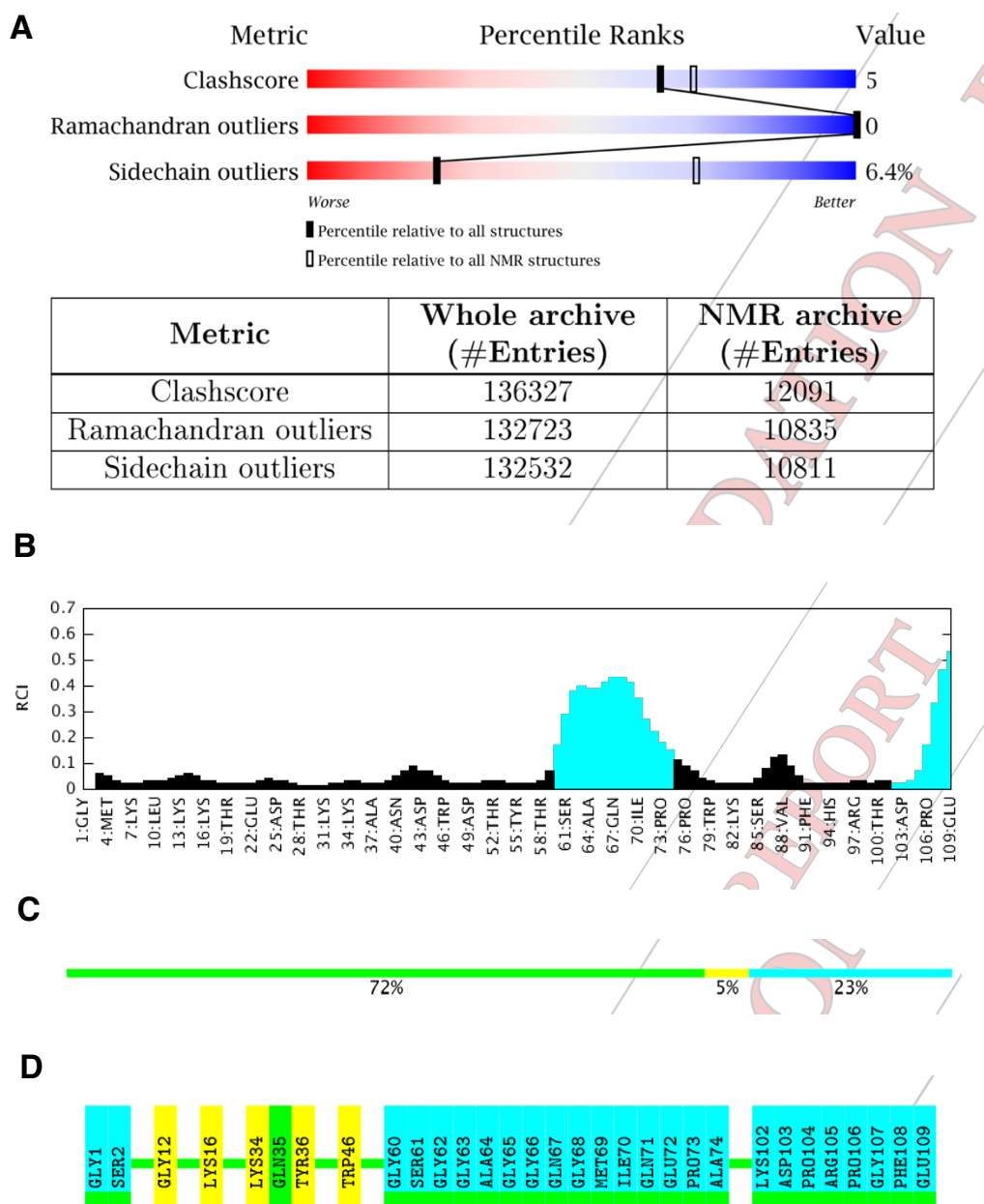


Figure 3.2.6 - wwPDB validation result summaries for refined GB1-WW4 ensemble.

A) Comparison of the final GB1-WW4 ensemble of 20 refined models with all/NMR only deposited PDB entries. Slider scale representing overall quality of the ensemble from worse to better than those in the database. B) Random coil index results for GB1-WW4 construct 1-109. The probability results for individual residues to occupy a random coil conformation based on the chemical shift data and amino acid sequence. Values higher than 0.2 indicate a significant likelihood of disorder. Black – well-defined residues; blue – poorly defined residues. C) Overall quality analysis of geometry across all 20 refined models for GB1-WW4. D) Individual residue geometric quality analysis for GB1-WW4 ensemble. Green - no outliers; yellow – one outlier; blue – poorly defined. Short green connecting segments represent 2 or more residues which have no outliers. Image derived from the preliminary report produced after wwPDB validation (validate-rcsb-3.wwpdb.org; Berman et al., 2003; Berman et al., 2007).

Criteria	GB1-WW4: 3-59 and 75-101 (ensemble average)
RMSD - backbone	
GB1: 3-59	0.26 Å
WW4: 75 - 101	0.3 Å
Bond lengths	
RMSZ	0.53 ± 0.11
#Z > 5	1 ± 2 / 693 (0.2 ± 0.2 %)
Bond angles	
RMSZ	0.50 ± 0.02
#Z > 5	0 ± 0 / 944 (0.0 ± 0.0 %)
Chirality outliers	0.0 ± 0.0
Planarity outliers	0.3 ± 0.05
Clashes	6 ± 2
Ramachandran analysis	
Analysed	84/109 (77 %)
Favoured	82 ± 1 (98 ± 1 %)
Allowed	2 ± 1 (2 ± 1 %)
Outliers	0 ± 0 (0 %)
% compared out PDB entries (NMR)	100 %
Side chain analysis	
Analysed	72/87 (83 %)
Rotameric	67 ± 2 (94 ± 2 %)
Outliers	5 ± 2 (6 ± 2 %)
% compared to PDB entries (NMR)	71 %

Table 3.2.3 - Summary of the criteria and validation results carried out by wwPDB on GB1-WW4 ensemble.

Model quality results from the preliminary wwPDB validation report on GB1-WW4 residues 3-59 and 75-101 (total = 84/109 residues). Results given are average values for across the structural ensemble. % in brackets where available. RMSD – root mean standard deviation; RMSZ – root mean square of all Z-scores. Adapted from the wwPDB validation report.

	Total	¹H	¹³C	¹⁵N
Backbone	410/416 (99 %)	164/166 (99 %)	166/168 (99 %)	80/82 (98 %)
Side chain	396/477 (83 %)	247/275 (90 %)	143/183 (78 %)	6/19 (32 %)
Aromatic	93/115 (81 %)	51/59 (86 %)	40/50 (80 %)	2/6 (33 %)
Overall	899/1008 (89 %)	462/500 (92 %)	349/401 (87 %)	88/107 (82 %)

Table 3.2.4 - Summary of assignment completeness for GB1-WW4 core regions as determined by wwPDB.

Data from the preliminary report provided by wwPDB validation service for core regions 3-59 and 75-101. 899 atoms out of 1008 for these core regions were assigned.

3.3. Discussion

Structure-based inhibitor design consists of several stages and rounds of modifications until a drug lead can move forward into clinical trials. Initially, the target of choice has to be cloned, expressed and purified. Next, structure determination is carried out to understand the spatial conformation of the protein. This develops into compound or peptide screening based on modelling of potential hits into favourable sites on the structure. Biochemical analysis then shortly follows to determine the affinity and efficacy of the compound/peptide. Without knowledge of the target structure, selective drug design towards a specific target is extremely difficult. In the case of WWP2, the presence of highly selective WW domains makes them ideal targets for novel anti-cancer therapies. Yet, none of the WW domains have been structurally solved to date, due to problems with protein stability and expression levels. The aim of this chapter was to reveal the structure of the fourth WW domain of WWP2 to facilitate the design of novel inhibitors against this E3 ligase.

The whole ensemble of structures performed well in the structure validation criteria. The results for GB1 alone, which as it is already published can act as a control, support the conformation shown in this chapter as the native form of WW4. Closer comparison of the published GB1 and the medoid model 16 (Figure 3.2.4B) showed that while a lot of the side-chains were very similar, there was still variations in these across the structure. This is probably due to the differences in construct/solution/environment between the two methods of structural solution and does not mean that those in this structural ensemble are necessarily incorrect. As for the backbone, there is little difference between the overall structures. The quality of the overall ensemble is supported by the absence of NOE violations above 0.3 Å, lack of dihedral outliers and low RMSD values. The presence of some errors in rotamer positions and aromatic bond lengths suggests some less than ideal aspects of the structure. Although, it is important to remember that comparison to only NMR-based structures, rather than to all PDB entries, the ensemble ranked better than average in quality. This may be due to more stringent validation parameters in place for X-ray crystal structures based on their fixed nature, though the vast majority are still applicable to NMR structures.

WW domain structures for other Nedd4 E3 ligases have previously been published, such as Nedd4 WW3 (PDB: 4N7F and 5AHT; Qi *et al.*, 2014; Panwalker *et al.*, 2016) and WWP1 WW4 domain (PDB: 2OP7), as well as Smad 7 bound structures of Smurf 2 WW2-3 (PDB: 2KXQ; Chong *et al.*, 2010) and Nedd4L (PDB: 2LTY; Aragon *et al.*,

2012) to name a few. Across these domains, a β -sheet consisting of three anti-parallel β -strands with a twist was observed, suggesting this conformation is a conserved folding pattern important to the function of these domains. Solution state NMR has revealed that the same is true for WW4 within WWP2. Collectively, they all show a twist in the first loop region towards the 2nd loop. This suggests that the curve may orientate the side chains of residues into favourable positions to facilitate substrate binding. Work carried out by Kasanov *et al.* (2001), with WW domains of proteins such as Nedd4, WWP1 and WWP2 found that the conserved histidine at position 94 and threonine at 99, according to WWP2 WW4 domain sequence shown in Figure 4.3.1A, were important in substrate recognition. When looking at these residues in WW4, as well as the XP groove consisting of tyrosine (Y90) and substituted phenylalanine (F101) (shown in Figure 4.3.1), the residues appear to reside on either side of the diagonal groove generated by this twist supporting the involvement of these residues in the function of this domain. Although the side-chain orientations of the residues H94 and F101 are not as well-defined as those located in the β -strands, potentially due to natural variation from solvent exposure in turn producing fewer restraints as a limited number of contacts are possible, they still present this overall pattern of defining a groove in the structure implicated in substrate recognition. Another important residue is valine 88 (V88), as this amino acid has been found to determine the class of the WW domain and in turn the substrate selection motif. Espanal and Sudol (1999) found that by mutating the amino acid at this position from a leucine to a tryptophan, the domain would change from a class I (PPxY) to a class II (PPLP). Mutation of the histidine residue at +2 had a lower degree of effect on motif recognition. This suggests that this being a valine in WW4, and likewise in the other three WW domains of WWP2, supports the recognition of these domains towards only PPxY-motif containing substrates.

A

WW4 : (444) 73-PALPPGWEMKYTSEGVRY**Y**FVD**H**NTRT**T****T****F**KDPRPGFE-109(477)

B

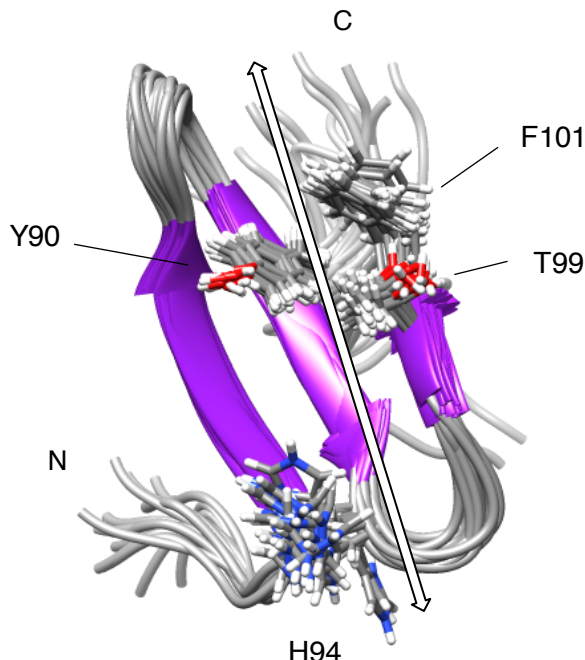


Figure 4.3.1 - WW4 structure showing XP groove and other conserved residues.

A) Amino acid sequence of WW4 with conserved residues and XP groove residues in bold. Number in brackets refers to numbering in WWP2-FL sequence. B) Side-chains of residues tyrosine 90 (Y90) and phenylalanine 101 (F101) that form the XP groove and conserved histidine 94 (H94) and threonine 99 (T99) for the 20 final structures of WW4. Arrow indicates potential region of peptide interactions.

Panwalker and colleagues (2017) found that the 3rd β -strand of Nedd4 WW3 was the first region to dissociate from the substrate as well as the one that significantly stabilised and increased rigidity upon substrate binding. As this region shows a very conserved structure and contains residues important for direct binding to the prolines of the PPxY motif, this may also be the case for WW4 in WWP2. This is further suggested from work carried out by Lloyd Wahl (2016) which identified the residues of the last β -strand as the ones which underwent the most perturbation in NMR titration analysis with substrate peptides. However, as other residues were also found to be involved in substrate binding in the other β -strands rather than just the conserved residues mentioned previously, this could imply an allosteric effect across the domain

rather than direct binding. Further analysis of the conformational changes upon substrate interaction is required to conclusively say whether this is the case for WW4. Nevertheless, designing peptides to mimic this partial flexibility of the 3rd β -strand alongside the other key residues may in fact be difficult, thus mimicking the substrate to competitively block the binding site may be a more appropriate avenue of investigation using structure-based methods.

4. WW domain interaction analysis

4.1. Introduction

The ubiquitination of substrates is important for the regulation of many cellular signalling pathways. Nedd4 family E3 ligases such as Nedd4, Smurfs 1 and 2, and WWP1 and 2 interact with secondary messengers such as PTEN and the Smad proteins involved in PI3K/AKT and TGF- β signalling respectively, as discussed in chapter 1. The overexpression of these ligases has been linked to oncogenesis via the increased degradation of these tumour suppressor proteins. WWP2 has been implicated in prostate cancer development, with significant influence from the WWP2-N isoform upregulated in earlier stages (Soond *et al.*, 2013), due to its ubiquitination of both PTEN and Smad proteins leading to PI3K/AKT and TGF- β overactivation and cellular survival. Thus, understanding these binding partners and how they interact with the different isoforms can direct the design of novel therapeutics, in which very few inhibitors are currently known. It is also of interest from a biological perspective to understand how the substrates are selected and targeted for degradation at a molecular level.

WWP2-FL was isolated in mass spectrometry analysis of PTEN interaction partners back in 2011 by Maddika and colleagues, introducing a new substrate to the few discovered for WWP2. PTEN is a negative regulator of the PI3K/AKT signalling pathway, described in detail in chapter 1. Overexpression and silencing of WWP2 showed a WWP2-dependence for the turnover of PTEN, contradicting previous data that Nedd4-1 was the primary E3 ligase that controlled PTEN turnover. This was supported by the lack of Nedd4-1 in the mass spectrometry and immunoprecipitation results (Maddika *et al.*, 2011), and later concluded that while Nedd4-1 does target PTEN, the dominant E3 ligase for PTEN ubiquitination is WWP2. Work by Chen *et al.* (2016) further supported this conclusion but also highlighted the importance of the PTEN phosphorylation status in controlling PTEN degradation by WWP2. The WWP2 interaction site on PTEN was isolated to between 100 and 187 residues, however there is no PPxY motif in this region, or anywhere in PTEN. This implies that either the WW domains of WWP2 are not exclusively group I WW domains and can bind to motifs other than PPxY, or that binding to PTEN is not mediated via the WW domains. There is also currently no published evidence of N and C isoform interactions with PTEN which may uncover how these two proteins interact and regulate activity levels in relation to PI3K/AKT signalling.

The independent folding domains of WWP2 have a layout common to the Nedd4 family of proteins consisting of an N-terminal phospholipid-binding C2 domain, a C-terminal domain containing the catalytic cysteine called the HECT domain, and four central WW domains. Soond and Chantry (2011) published data confirming the presence of WWP2 isoforms with WW domains from opposing regions of the full-length protein, as discussed in chapter 1. Immunoprecipitation assays revealed that WWP2-FL was capable of binding Smad proteins 2, 3 and 7 but had preferential binding to Smad 7. In comparison, WWP2-N only interacted with R-Smads 2 and 3, whereas WWP2-C bound to Smad 7 only. WWP2-FL interactions with Smad 7 were also further increased by TGF- β stimulation. Interestingly, WWP2-N was also shown to interact with WWP2-FL. They proposed that in the absence of TGF- β stimulation, WWP2-FL remains in an autoinhibitory conformation, evidence of which was later published by Chen *et al.* (2017) establishing the 2-3 linker as the mediator of this autoinhibition. However, when WWP2-N is expressed under TGF- β negative conditions, this isoform interacts with and activates WWP2-FL leading to rapid degradation of regulatory Smads 2 and 3. Dimerization of WW domains has previously been found for Smurf 1/2 (Aragón *et al.*, 2012) which may be how these two isoforms interact to relieve the autoinhibitory conformation. This results in activated WWP2-FL which is primed for substrate interactions. In turn, when the pathway is stimulated, the preferred target for WWP2-FL changes towards inhibitory Smad 7 which is also targeted by the C isoform, although the regulation of isoform expression is still unclear. Due to WWP2-N lacking any intrinsic ubiquitination activity, its interaction with Smads 2 and 3 is thought to stabilise the expression levels of these proteins by competing with WWP2-FL, or by recruiting these proteins to WWP2-FL for ubiquitination (Soond and Chantry, 2011). Understanding the nature of these interactions may reveal more about this bias in substrate targeting by the isoforms and lead to ways of selectively targeting these interaction patterns.

The structural basis of the substrate interactions with Nedd4 family proteins are not fully known due to difficulties obtaining stable and abundant independent WW domains for structural studies. The WW domains of WWP2 are 33-34 amino acids in length containing 1-2 tryptophan residues, fold into the characteristic three antiparallel β -strands, similar to the structure shown in the previous chapter for WW4, and commonly bind to PPxY motifs. Sequence alignment between each of the WW domains (Figure 4.1.1) highlighted WW1, 2 and 3 as 70-75 % similar, however comparison of 1, 2 and 3 each with WW4 was between 55-60 %. This suggests the difference in binding

preferences between WW1 and WW4 as seen via the N and C isoforms is not particularly surprising. However, little is currently known about the binding activities of WW2 and WW3 within WWP2-FL, which may help elucidate whether this sequence similarity is reflected in binding preferences and affinities, or whether it is the structure and side chain rotamer orientations that decide this.

WW4	P	A	L	P	P	G	W	E	M	K	Y	T	S	E	G	V	R	Y	F	V	D	H	N	T	R	T	T	T	F	K	D	P	R	P
WW3	G	P	L	P	P	G	W	E	K	R	Q	D	-	N	G	R	V	Y	Y	V	N	H	N	T	R	T	T	Q	W	E	D	P	R	T
WW1	D	A	L	P	A	G	W	E	Q	R	E	L	P	N	G	R	V	Y	Y	V	D	H	N	T	K	T	T	T	W	E	R	P	L	P
WW2	R	P	L	P	P	G	W	E	K	R	T	D	P	R	G	R	F	Y	Y	V	D	H	N	T	R	T	T	T	W	Q	R	P	T	A
			*	*		*	*	*	:					.	*		*	:	*	:	*	*	*	:	*	*	:	*	:	:	*			

Figure 4.1.1 - Sequence comparison of WWP2 WW domains.

Alignment analysis of the amino acid sequences for the four WW domains present in WWP2. * refers to fully conserved residues; : represents conservation of residue properties of strong similarity; . represents conservation of residue properties of weak similarity.

There is building evidence of WW domains acting in tandem rather than exclusively independent, with 5 proposed modes of action (Dodson *et al.*, 2015). By acting in pairs, this enables a tighter regulation of substrate selection and ubiquitination in a cell-based context. Evidence of increased binding affinity by the interaction of both WW domains to different motifs where the binding of one is regulated by the binding of the other has been found for Nedd4L (Aragón *et al.*, 2011), and suggested for Smurf 1/2 (Chong *et al.*, 2010). Nedd4L WW2 and WW3 domains interacted with phosphorylated Smad 3 at adjacent binding motifs. WW2 recognised the PPxY motif and a phospho-Threonine (pT) site whereas WW3 bound to phospho-Serine (pS) adjacent to the other motifs, with the binding of WW3 leading to an increase in overall affinity. Similarly, Smurf 2 showed cooperative WW domain binding to Smad 7 via WW2 and WW3 where WW2 created additional contacts with negatively charged residues and prolines N-terminal to the PPxY motif to increase the binding of WW3 to Smad 7. However, this is controversial as data published by the same group a year later suggested that rather than interacting with an upstream motif to increase binding, it might be that the ability of WW2 to dimerise results in increased stability of the tandem domain leading to a rise in affinity (Aragón *et al.*, 2012). While both of these examples involving Nedd4 E3 ligases highlight the occurrence of 2 of 5 models in tandem domain cooperativity (Dodson *et al.*, 2015), it does not rule out the possibility of independent binding to different motifs leading to an additive affinity, a change in ligand preference of the first domain due to the presence of the second altering the structure of the first, or the

presence of the second domain reducing the likelihood of interaction by obtaining an unfavourable conformation. Another aspect discovered to influence tandem domain activity was the length of the linker region between the two domains. A reduced affinity for Smad 7 was observed for the longer Smurf 1 isoform, which has an extra 26 residues, unlike the shorter isoform which used WW domain coupling in its interaction (Chong *et al.*, 2010). The lack of linker between WW1-2 in WWP2 and only 6 amino acids between WW3-4 suggests that WW domain coupling may be a regulatory mechanism adopted by WWP2-FL for controlling substrate interactions.

Changes to the protein-protein interactions discussed above by overexpression of one or more component/isoform contributes to the development of cancer. Targeting these protein-protein interactions has become increasingly popular for drug design. Peptide-based therapies as novel drugs are favourable over small molecules due to their specificity and increased efficacy towards the intended target. Evidence from the analysis of protein-protein interfaces showed an increased propensity for certain amino acids such as aromatic residues (Sillerud and Larson, 2005). It is thought that these residues along with arginine and a few aliphatic residues (proline and isoleucine) which are predominant at interaction sites contribute to the overall binding energy of the interaction. Additionally, the presence of 2 tryptophan residues (phenylalanine in the place of the second in a few cases) and other favourable residues in these WW domains make them ideal templates. This is why understanding which residues are involved in WWP2: substrate binding will aid the design of peptidomimetics against the interactions associated with oncogenesis.

The use of NMR titration analysis can identify the key residues involved in protein-protein interactions. NMR titrations reveal changes in the protein upon substrate binding at an atomic level. [¹H-¹⁵N]-HSQCs are used in this project to trace the chemical shift changes of amide groups induced by either direct binding of the peptides or by conformational alterations induced by the binding. A change in the chemical environment of an atom, referred to as exchange, is reflected in an alteration in the spin system resonant frequency (Keeler, 2010), which can be visibly observed in the spectrum after data is collated. Three possible types of exchange may occur during a protein-ligand interaction experiment determined by the rate constant for exchange. Each nucleus can exist at two chemical shift values, one characteristic of the unbound conformation (A) and the other the bound conformation (B). If the interaction results in many exchanges between the unbound and bound conformations during the

experimental time-frame then an average of the two chemical shifts is observed. This is referred to as fast exchange. The observed chemical shift value is dictated by the proportion of nuclei in each conformation. For example, at the start of a titration more nuclei precess at the frequency of the unbound conformation, producing a peak at chemical shift A. As ligand is gradually added, the population may shift towards the bound conformation resulting in steady movement towards chemical shift B. In the absence of exchange within the detection period (slow exchange), if only one conformation exists then only one peak is observed, or if there are populations of both conformations, peaks are observed at positions A and B. The last possible exchange type is intermediate exchange. This is where the exchange rate between the conformations increases enough that the time of precession at each frequency ultimately cancels each other out and results in line broadening. Understanding each of these situations can provide information about key residues on the interaction interface. As such, peaks which are involved in peptide binding or undergo allosteric changes will migrate towards the bound chemical shift position according to the rate of binding. The association between the rate of exchange and ligand binding can be used to calculate dissociation constants for the specific residues involved in the interaction. In some cases, however, peaks can disappear and reappear at a different chemical shift making it difficult to determine a dissociation constant. The focus of this work however, was using NMR titrations to further understand the key residues involved in the interactions rather than affinities.

In the previous chapter, a structure-based approach to targeting the overexpression of WWP2 was discussed in relation to the fourth WW domain. Building upon this concept, this chapter discusses WWP2 interactions with substrates: PTEN, Smad 2, 3, and 7, directed towards the use of NMR titrations in peptidomimetic design. Given that PTEN lacks a known WW domain interaction motif, the initial aim was to determine whether the three WWP2 isoforms FL, N and C could interact with PTEN. The next aim was to use titration analysis to understand the binding of substrates on an atomic level to WWP2. Initially, the interaction with PTEN was the focus, however, due to problems isolating a single region of WWP2 that interacted with PTEN, the aim changed to look at Smad protein binding to WW1 due to its presence in the N isoform and the WW1-2 tandem domain which lacks an obvious linker region, using peptides based around the PPxY motifs. The objective was to identify the key residues involved in substrate binding in order to design peptidomimetics, which with the presence of two tryptophan

residues and several aromatic groups provided a favourable basis for inhibitory peptide design.

4.2. Results

4.2.1. WWP2-PTEN interaction analysis

Previous research indicated that PTEN interaction with WWP2 is mediated via amino acids 100-187 of PTEN (Maddika *et al.*, 2011). However, the region within WWP2 that binds to PTEN remains unknown. Due to the lack of PPxY motif in PTEN, this also suggests that the WW domains of WWP2 may interact with alternative substrate-based motifs. Therefore, pull-down analysis of WWP2 isoforms with PTEN may provide more information about how WWP2 targets this substrate. In order to look at the binding of WWP2 isoforms to PTEN, full-length PTEN (GST-PTEN-FL), and GST-PTEN deletion constructs 1-274 and Δ 274-342 (Figure 4.2.1A) were bacterially expressed and purified on glutathione beads. WWP2-FL, WWP2-C and WWP2-N were expressed as Flag-tagged proteins in mammalian HEK 293A cells, and the lysate added directly to the bead samples. Analysis was carried out on the proteins bound to the beads using SDS-PAGE and western blotting with anti-flag antibody. Results showed that WWP2-FL bound to full-length PTEN but not constructs 1-274 and Δ 274-342 (Figure 4.2.1B), potentially due to lower affinity for mutated forms of PTEN perhaps due to loss of an additive binding region although this requires further investigation. In contrast to this, WWP2-C showed binding to only PTEN-FL and Δ 274-342, whereas WWP2-N interacted with all constructs. This suggested that the WWP2 isoforms may interact differently with PTEN, potentially resulting in different consequences. It is important to note that the majority of expressed GST-PTEN constructs were present in the insoluble material during purification, suggesting the presence of inclusion bodies. This may explain the difference in WWP2-FL results as the abundance of constructs 1-274 and Δ 274-342 was lower than GST-PTEN-FL as seen in the SDS-PAGE image for the samples used (Figure 4.2.1B lower panel), however the results of N and C isoform pull-downs would indicate otherwise.

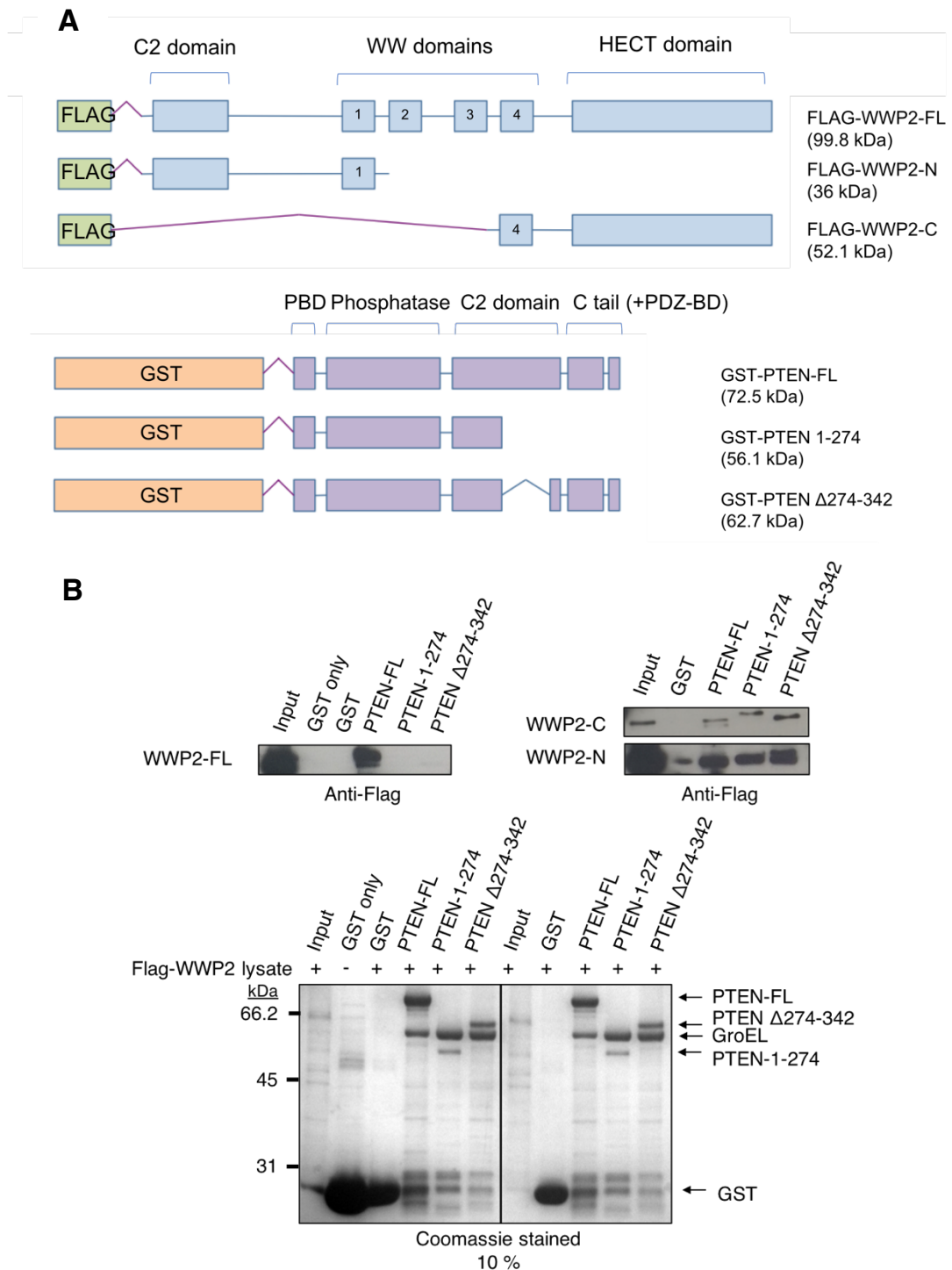


Figure 4.2.1 - WWP2 isoform interaction analysis shows binding differences.

A) Schematic representation of the constructs used in B. Flag-tagged isoforms: WWP2-FL (full-length), -N (N-terminal) and C (C-terminal); GST-tagged tumour derived PTEN constructs spanning amino acids 1-403 (FL - full-length), 1-274 and deletion construct missing residues 274-342. GST – glutathione S transferase (tag); PBD – phospholipid binding domain; PDZ-BD – PDZ binding domain. B) GST-coated or GST-PTEN-coated glutathione beads were incubated with cellular lysate containing expressed Flag-WWP2-FL, -WWP2-C or -WWP2-N. Beads were

then washed, and the proteins eluted. Samples were analysed by SDS-PAGE and western blot using anti-flag antibody (n=1). Negative controls included 'GST only' - GST-coated beads incubated in buffer, and 'GST' - GST-coated beads incubated with lysate expressing flag-tagged protein. GroEL refers to the chaperone protein present in bacteria such as *E.coli*.

To further investigate the difference in isoform binding, PTEN deletion constructs were synthesised by mutagenesis spanning from the first amino acid up to the end of the phosphatase domain (1-185) or to the end of the C2 domain (1-350) (Figure 4.2.2A). Constructs were expressed and purified on glutathione beads as described in section 2.2.4.5. Pull-down analysis using lysates expressing the N and C isoforms was carried out and the purified proteins analysed by SDS-PAGE and western blot. The N-isoform is known to run at ~50 kDa and the C-isoform at ~45 kDa (Wahl, 2016), which was further backed up in this work by plasmid sequencing to verify the plasmids and associated expression bands on SDS-PAGE. A band consistent with the N isoform is present in all PTEN sample lanes suggesting binding occurs between each N-terminal domain of the proteins (Figure 4.2.2B). In contrast, the C isoform showed binding to full-length PTEN and no interaction with 1-350 truncated protein. It is unclear whether the C isoform interacts with construct 1-185 as both proteins migrate to ~45 kDa. The presence of PTEN bands on the anti-Flag western blot, in particular the WWP2-N + PTEN 1-185 lane, suggested that this band may be due to non-specific antibody interaction with GST-PTEN 1-185 and not the C isoform. However, this cannot be concluded and would require further analysis. Overall the data supports the conclusion that the isoforms interact with alternate ends of the PTEN protein, perhaps indicating separate mechanisms of regulation and turnover dependent on the isoform and cellular signalling status of PTEN activity.

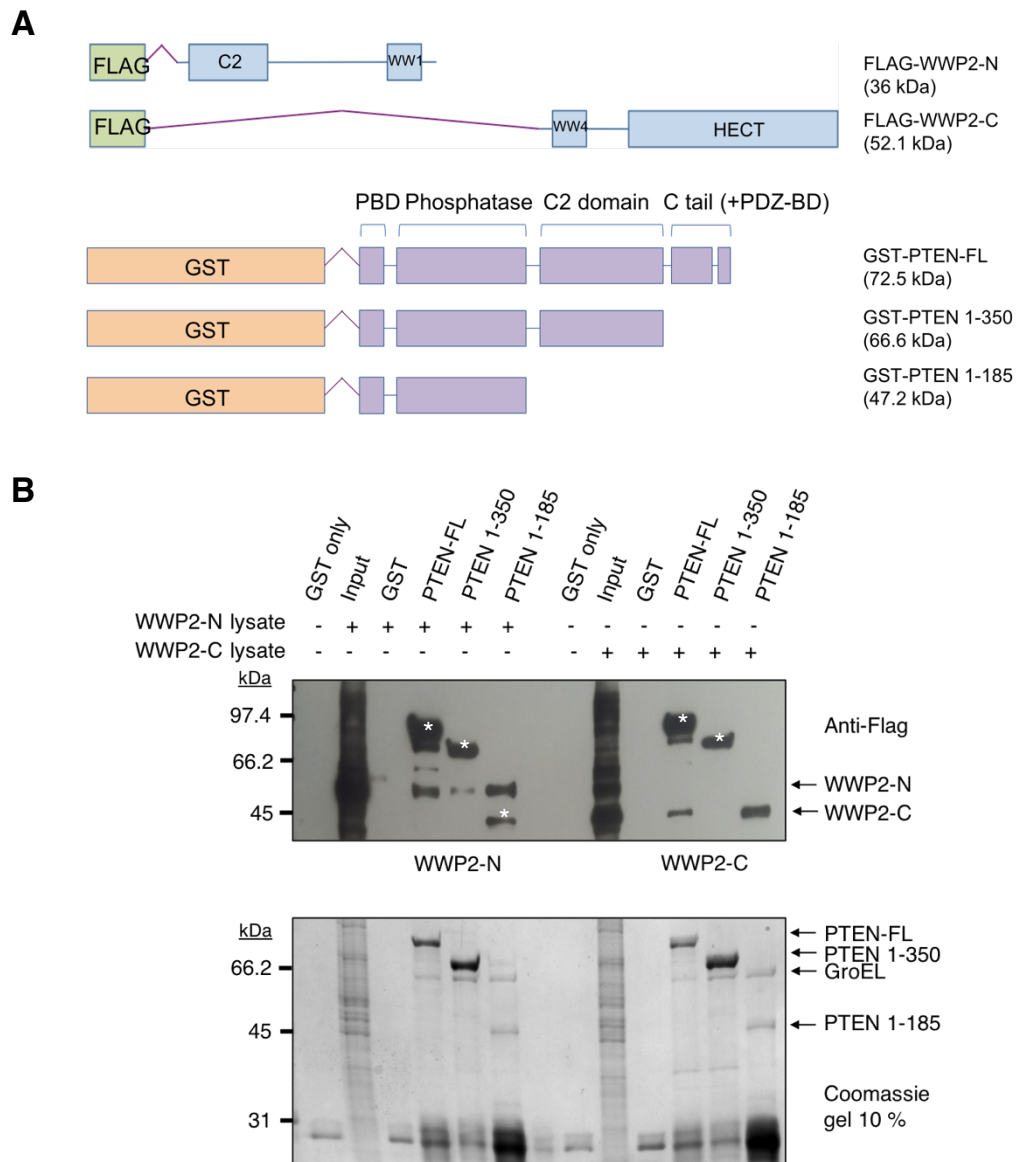


Figure 4.2.2 - Pull-down analysis of PTEN deletion construct interactions with WWP2 isoforms N and C.

A) Schematic diagram showing the constructs of GST-PTEN and Flag-WWP2 used in part B. PBD - phospholipid binding domain; PDZ-BD - PDZ binding domain; FL – full length. B) Glutathione beads bound by GST or GST-PTEN constructs were incubated with flag-tagged N and C WWP2 isoform containing lysates. Bound proteins were eluted and analysed by SDS-PAGE and western blot against flag-tag (n=1). Negative controls were GST-coated beads either incubated in buffer ('GST only' lane) or with isoform lysate ('GST' lane). White asterisks (*) refer to PTEN bands based on molecular weight migration pattern.

Another aspect of interest are the WW domains, as these convey substrate specificity. Due to the lack of PPxY motif in PTEN, the next step was to determine whether PTEN is capable of interacting with WW domains, as the presence of WW1 in the N isoform and WW4 in the C isoform may be involved in PTEN binding. Due to difficulties expressing individual WW domains for pull-down analysis, HA-tagged tandem domain constructs WW1-2 and WW3-4 were used (produced by Tiffany Yim). Constructs were expressed in HEK 293A cells and the lysates added to GST-PTEN-FL, GST-PTEN 1-185 or GST-PTEN 1-100 coated beads. Constructs are outlined in Figure 4.2.3A. After incubation, the beads were washed, and purified proteins analysed by SDS-PAGE and western blot. Results showed that WW1-2 only interacted with full-length PTEN, whereas WW3-4 showed interaction with all constructs (Figure 4.2.3B). These results suggest that these tandem domains also interact with opposing sides of PTEN. Compared to the isoform results, these tandem domains showed a different pattern of binding, suggesting that perhaps when expressed as isolated regions, these domains seem to be able to bind to alternative sites in PTEN, perhaps one that is adjacent to the N-isoform C2 interaction site or HECT binding site for the C isoform in PTEN. However, it was interesting that the WW domains were still able to interact with these constructs despite the lack of the well-known motif targeted by these WW domains and further investigation may reveal the amino acids involved in the interaction within both proteins. Collectively, the data suggests that the binding region of WWP2 to PTEN seems to depend on the isoform and that the binding site may in fact consist of several regions across the WWP2 protein and PTEN. For this reason, along with the low bacterial expression levels of PTEN constructs due to the formation of inclusion bodies, further investigation of this interaction using NMR was not pursued.

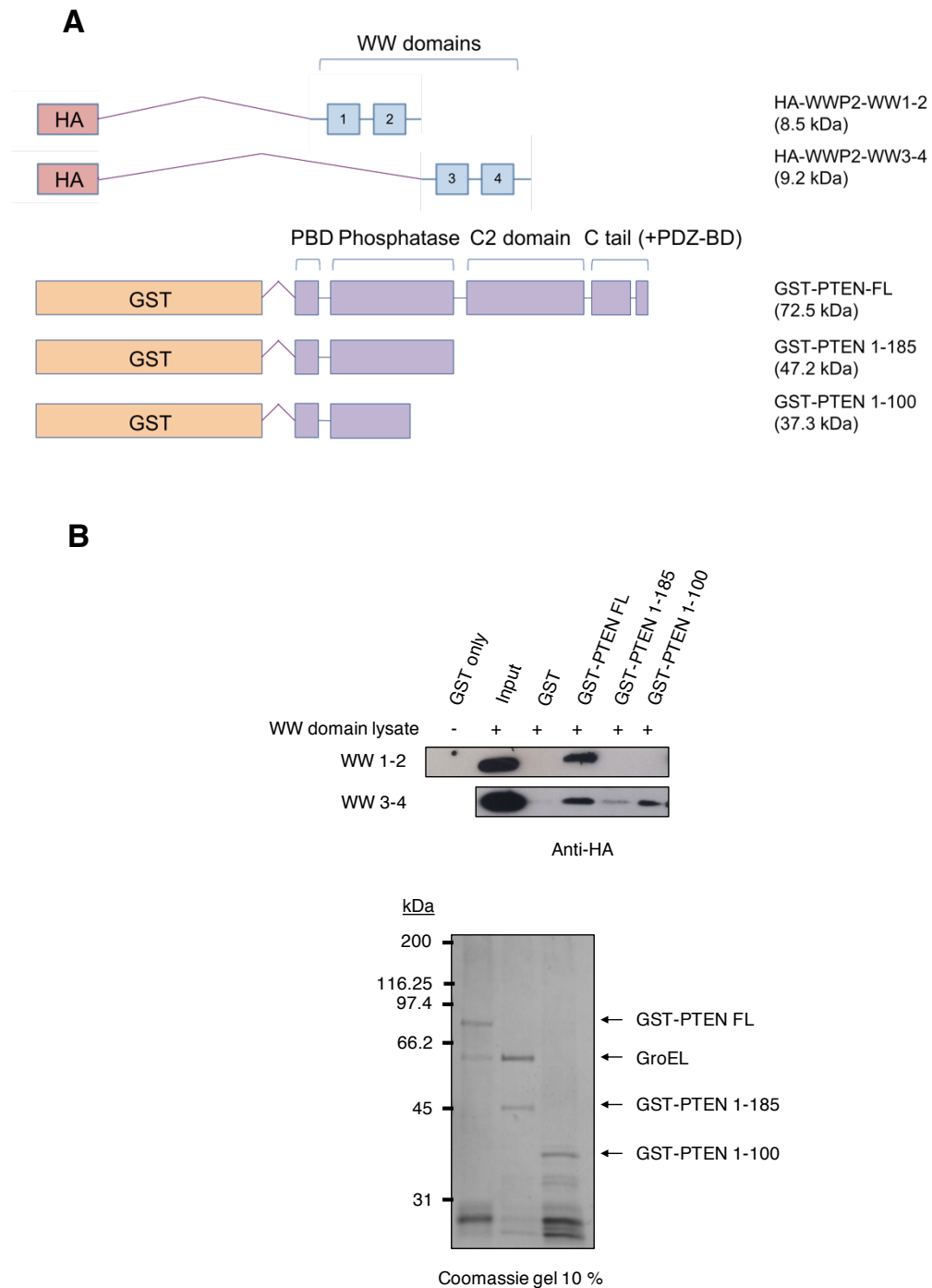


Figure 4.2.3 - GST-pull-down analysis of tandem WW domains with PTEN constructs.

A) Representative diagrams of the constructs used in part B. WW – double tryptophan domain; HECT – homologous to E6AP C-terminal domain; HA – tag derived from human influenza hemagglutinin; GST – glutathione S transferase (tag); PBD – phospholipid binding domain; PDZ-BD – PDZ-binding domain B) GST-coated beads or GST-PTEN-coated beads were incubated with lysates expressing HA-WW1-2 or HA-WW3-4 then washed several times. Bound proteins were analysed by western blotting against HA (upper panel). SDS-PAGE was used to determine the levels of PTEN in the experiment (lower panel). Images in B representative of 2 experiments (n=2).

4.2.2. Smad peptide generation

Due to pull-down assays indicating the binding region between PTEN and WWP2 may actually include several regions across WWP2, this meant analysis of the PTEN-WWP2 interaction was not investigated by NMR titrations at this stage. Furthermore, the presence of inclusion bodies reducing the proportion of soluble PTEN constructs suggested that obtaining a high enough concentration would also have proven challenging. Therefore, focus shifted towards the Smad proteins 2, 3, and 7 which are known interactors of WWP2 WW domains (Soond and Chantry, 2011).

In order to carry out titration analysis, peptides were designed based on the PPxY motif in the linker regions of Smad 2 and 3 and incorporating residues on either side of this motif, outlined in Figure 4.2.4A. Bacterial expression and purification was investigated for the production of recombinant peptides. Previous work in the lab showed that this was possible for Smad 7 peptide but had yet to be tested for Smad 2 and 3. pET21d plasmids containing the sequences for Smad 2 and Smad 3 peptides were generated using in-fusion cloning to create His-SUMO-fusion constructs. Bacterial expression and Ni-affinity purification showed high quantities of recombinant protein could be expressed and isolated (Figure 4.2.4B). However, upon digestion using commercial SUMO-protease at room temperature after dialysis into SUMO-protease buffer, as recommended, no digestion was present in the first 21 hours (data not shown). After further incubation, precipitate formed before complete digestion could be achieved. Analysis of the precipitate showed increasing MW bands from 6.5 kDa all the way up to 200 kDa, as well as undigested and digested SUMO construct bands (Figure 4.2.5A). This suggested partial digestion and potential aggregation of the free peptides. A sample of in-lab expressed and purified ULP-1 was also tested at a ratio of 0.5:60 to determine whether the previous sample had low activity levels or whether the added quantity could increase the digestion rate. No precipitate formed with the new UPL-1 present, however by 72 hours at room temperature degradation was observed with incomplete digestion at up to 50 % which did not occur for Smad 7 peptide generation, shown in Figure 4.2.5B. Taking forward the partial digestion to see whether any peptides were present, results showed faint peptide bands at 2.5 and 3 kDa equating to the size of the Smad 2 and Smad 3 peptides, respectively. This suggests that the generation of these peptides may be possible, however when this was repeated, using ULP-1 at a 0.25:30 ratio, degradation was observed before adequate digestion (data

not shown). It is possible that the protein molecules were folded in a manner which sterically hindered access to the di-Gly digestion site in the majority of the sample, or that the original constructs are just too unstable for the timeframe required for complete digestion. The latter is suggested due to the lower MW bands present in some fractions at the initial purification stage (Figure 4.2.4B), although this may be residual non-specific proteins from the Ni-affinity column. One aspect which could have had an effect was the concentration of the protein sample, as the expression level was very high and the volume at digestion was only 25-30 mL. Overall, Smad 2 and 3 peptide generation using bacterial expression and purification was unsuccessful at this stage, but it may be possible with further optimisation analysis. Therefore, synthetic peptides with the sequences described in Figure 4.2.6 were purchased for use in titration analysis.

A

SMAD 2: His-SUMO- (210) IEPQSNYIPETPPPGYISEGE (231)

SMAD 3: His-SUMO- (170) IEPQSN-IPETPPPGYLSEGE (190)

B

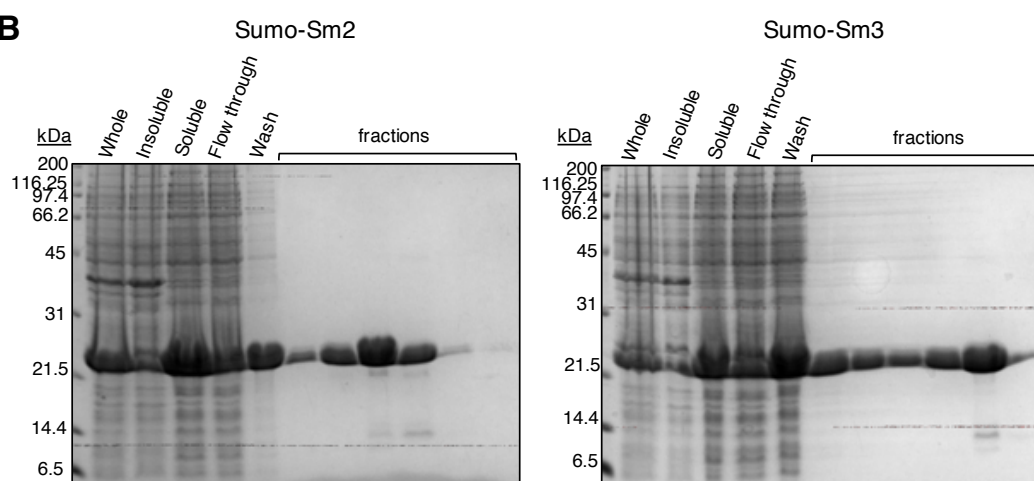


Figure 4.2.4 - His-SUMO-Smad 2 and -Smad 3 peptide construct purification by Ni-NTA affinity purification.

A) Amino acid sequences for the Smad 2 and Smad 3 peptide constructs. B) 15 % SDS-PAGE analysis of the whole, insoluble, soluble, and Ni-affinity purifications of Smad 2 and Smad 3 constructs. Purification was performed using ÄKTA-FPLC with fractions collected under UV monitoring at 280nm, with every other the peak fraction analysed by SDS-PAGE. Representative gels for 4 experiments (n=4).

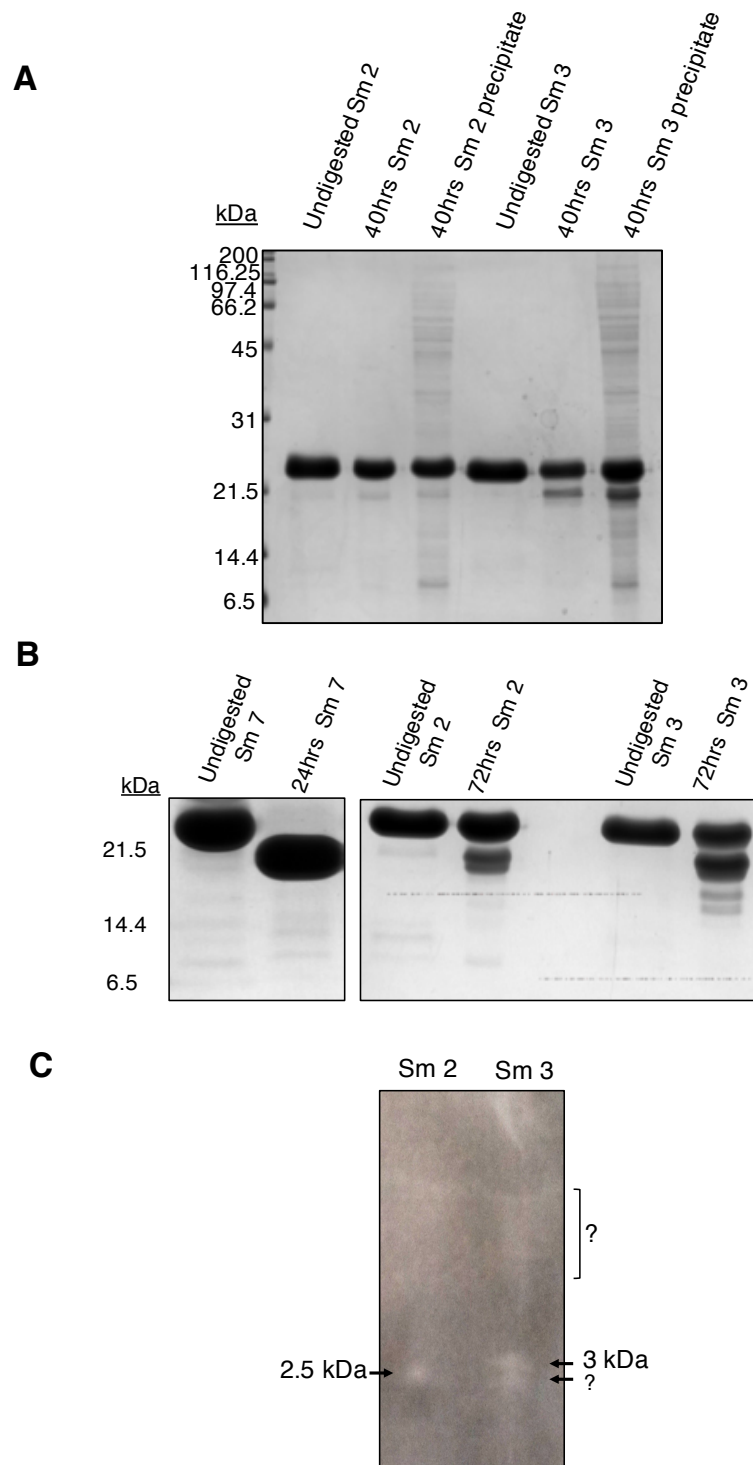


Figure 4.2.5 - Digestion analysis of Smad 2 and Smad 3 peptide constructs.

A) SDS-PAGE analysis using Coomassie staining of digestion samples with commercial SUMO protease. B) Comparison of His-SUMO-Smad 7 peptide digestion to His-SUMO-Smad 2 and -3 using in-lab expressed ULP-1 at a ratio of 0.25:30 (ULP-1: His-SUMO-Smad2/3). C) Silver stained tricine gradient gel of Smad 2 and Smad 3 peptide samples. ? indicates potential contaminant bands. Gel images representative from 4 experiments (n=4).

SMAD 2: (217) IPETPPPGYISEDGE (231)

SMAD 3: (176) IPETPPPGYLSEGE (190)

SMAD 7: (203) ELESPPPPYSRYPMD (217)

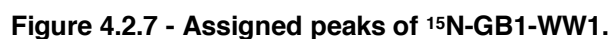
Figure 4.2.6 - Synthetic peptide sequences for Smad 2, 3 and 7 used in titration analysis.

Amino acid sequences for peptides based on the PPxY motif and surrounding residues of Smad 2, Smad 3 and Smad 7, with the start and end residue numbers according to the full-length proteins.

4.2.3. GB1-WW1 backbone assignment

Peaks in NMR spectra can be allocated to the specific nuclei in the protein residues based on the chemical shift positions, often referred to as peak assignment. The assignment of amide groups in [¹H-¹⁵N]-HSQC spectra for each residue is necessary for understanding NMR titration data and is the starting point for NMR-based structural analysis of proteins. To investigate the WW1 domain of WWP2, the domain was generated as a fusion protein with GB1 to overcome previous insolubility problems seen with WW domains when expressed on their own (Kristine Kay, 2012, personal communication). His-GB1-WW1 was bacterially expressed in double-labelled ¹³C-¹⁵N minimal media, digested using thrombin and purified using affinity-tag purification. HSQC, HNCaCb and CbCaCONH NMR spectra were collected at 800 MHz. GB1 assignments were obtained from Lloyd Wahl. No significant difference in peak positions were observed between GB1 when fused to either WW1, WW3 or WW4 (data not shown) suggesting GB1 has no significant interaction with the WW domain, and only influences the solubility of the construct by preventing aggregation. Sequential assignment of ¹³C-¹⁵N-GB1-WW1 was carried out using CcpNmr Analysis. Overall, 86.5 % of GB1-WW1 was assigned (70 % of WW1 only) with the exception of proline residues and 1 Gly, 2 Ser, 63 Ala, 83 Tyr, 84 Val, 85 Asp, 86 His, and 95 Arg. A [¹H-¹⁵N]-HSQC spectrum displaying the assigned peaks and 10 unassigned peaks is shown in Figure 4.2.7A. The presence of peak broadening, shown by residues 50 Asp, 70 Gly and 79 Gly for example (Figure 4.2.7B), suggests that these residues are undergoing slow chemical exchange in the experimental timeframe, potentially due to more than one conformation. The unassigned peaks may therefore be from an alternative conformation of the protein, rather than the missing peaks for the residues mentioned above, as suggested by the C α and C β chemical shift values (data not shown). Also, the clustering of many WW1 peaks towards the centre of the spectrum,

A



108

4.2.4. Titration analysis of WW1 interactions with Smad protein-based peptides

Evidence published by Soond and colleagues (2013) highlighted the preferential binding of the WW domains in WWP2 towards particular Smad proteins. Synthetic peptides based on the Smad protein PPxY region were used to investigate the binding between WW domains and Smad proteins. The following work looks at WW1 interactions using NMR titration analysis.

4.2.4.1. GB1-WW1 titration with Smad 2 peptide

To investigate the interaction of Smad 2 with WW1, purified ^{15}N -GB1-WW1 in NMR buffer was prepared as described in section 2.4.1. A 1D- ^1H -NMR spectrum and [^1H - ^{15}N]-HSQC (800 MHz) were obtained before starting the titration to check the sample for degradation or contaminants and provide a pre-titration spectrum for comparison to titration points with the presence of peptide. A few differences in the central cluster of peaks in the HSQC were observed (black boxes in Figure 4.2.8) compared to the previous HSQC, suggesting slightly different chemical environments and motions than the previous preparation of ^{15}N -GB1-WW1, however as the shifts and number of new peaks are minimal, this was not deemed a problem for further work. The titration was carried out as described in section 2.4.3 with sequential additions of synthetic Smad 2 peptide up to a ratio of 1:0.92. Due to loss of the sample, no further titration points could be carried out. Very few changes were observed as shown in Figure 4.2.9, which includes the increasing peak intensity of 87 Asn, decreasing intensity and peak migration for 92 Thr (Figure 4.2.9B), and the appearance of an unassigned peak. These changes could be explained by peptide interaction with these residues inducing a change in folding, but as 92 Thr presents with exchange broadening with no peak reappearance, this implies incomplete binding, which is explainable by the lack of peptide saturation. As the titration was only to a 1:0.92 ratio with incomplete binding, due to time restrictions preventing repeat expression and purification of ^{15}N -GB1-WW1, further titration points are required to conclusively determine the interacting residues of WW1 with Smad 2 peptide.

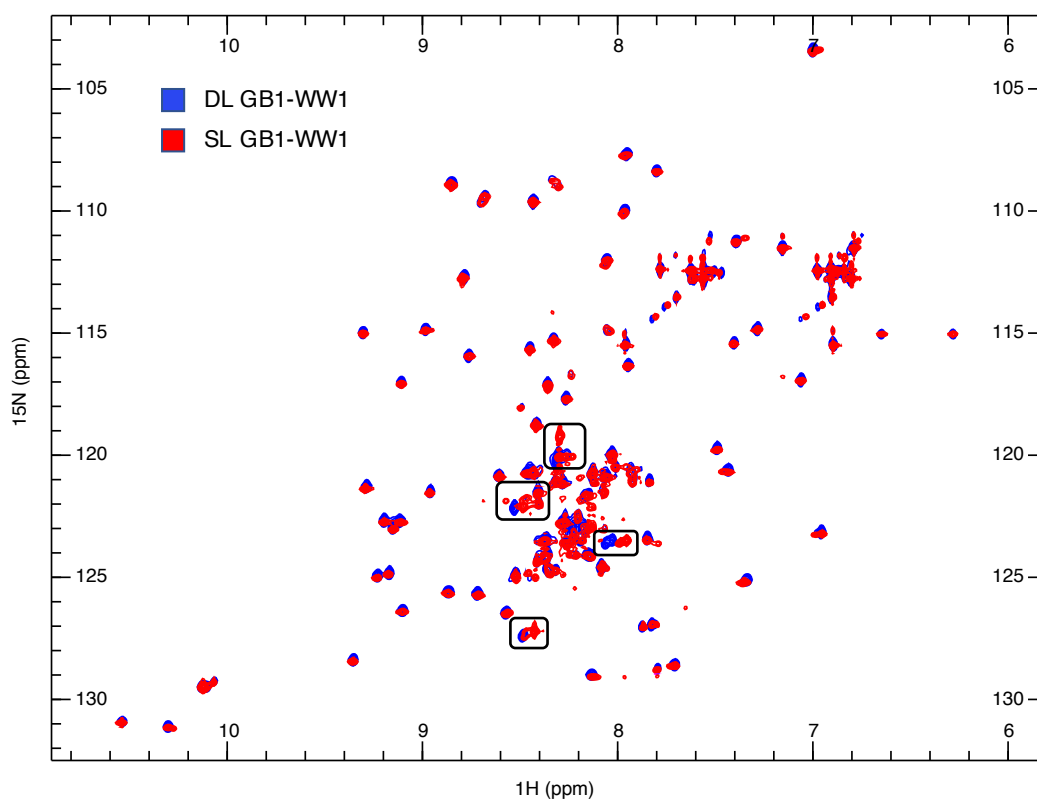


Figure 4.2.8 - [^1H - ^{15}N]-HSQC comparison of a ^{13}C - ^{15}N -GB1-WW1 sample and ^{15}N -GB1-WW1 sample.

Overlaid HSQCs of double-labelled (DL) GB1-WW1 used for backbone amide assignment (blue) and single-labelled (SL) GB1-WW1 used for the Smad 2 titration (red). ppm – parts per million.

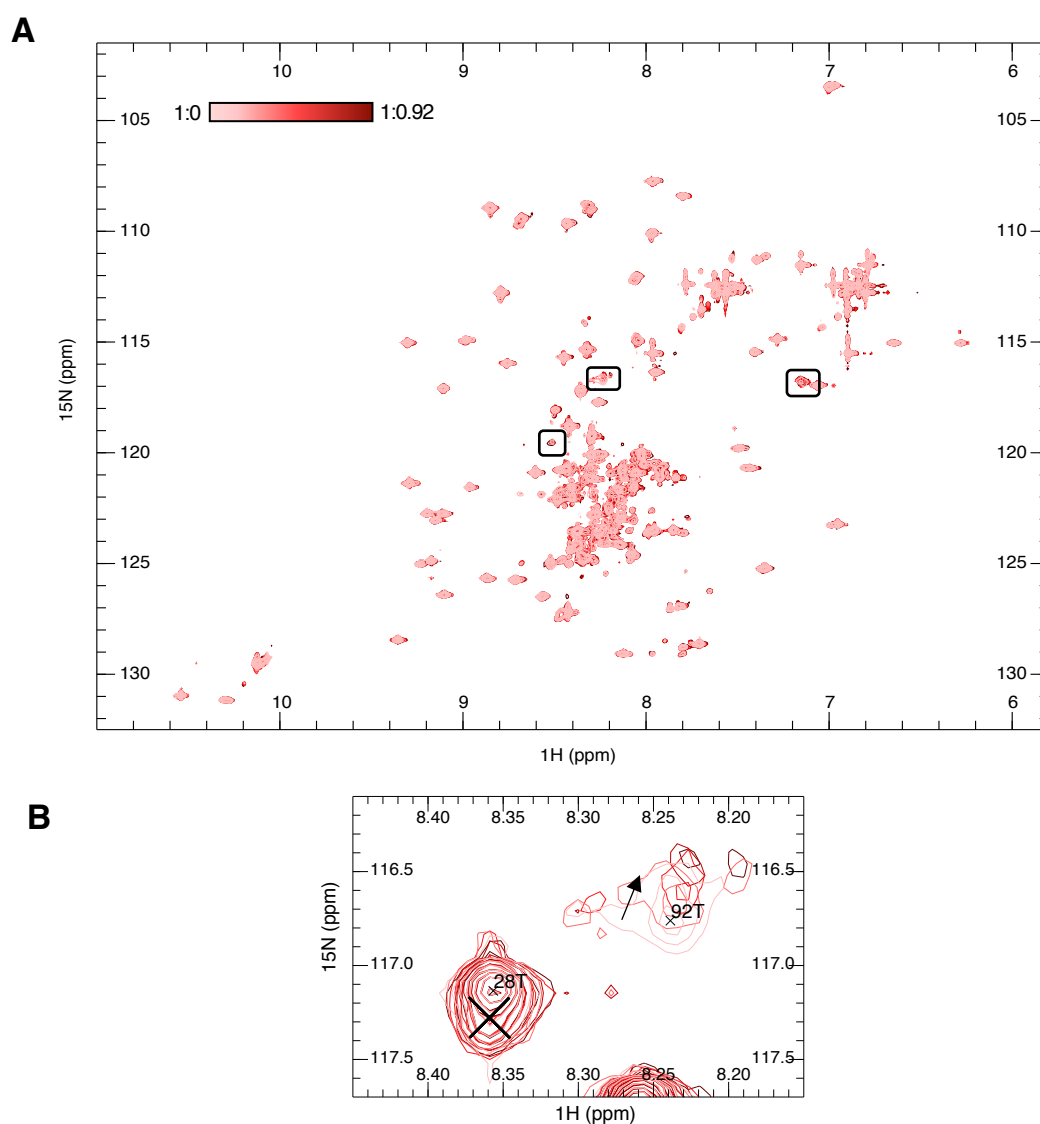


Figure 4.2.9 - Titration of GB1-WW1 with Smad 2 peptide.

Synthetic Smad 2 peptide was added sequentially to a sample of ^{15}N -GB1-WW1, ranging from a ratio of 1:0 (light pink) to 1:0.92 (red). A) Overlaid $[^1\text{H}\text{-}^{15}\text{N}]$ -HSQCs for ratio points 1:0 to 1:0.92. B) Spectral region showing decreasing intensity and peak shift for residue 92 Thr (92T), alongside an unmoving GB1 peak for comparison. X indicates no peak shifting, whereas the arrow indicates movement. The concentration of ^{15}N -GB1-WW1 was 0.6 mM across all titration points, with a final Smad 2 peptide concentration of 0.55 mM. ppm – parts per million.

4.2.4.2. GB1-WW1 titration with Smad 3 peptide

WWP2-N isoform has been shown to interact with both Smad 2 and Smad 3 so the next step was to investigate the interaction of WW1 with Smad 3 peptide. A titration using synthetic Smad 3 peptide was carried out as described in section 2.4.3 at 500 MHz. Several peaks were shifted by the end of the titration as shown in Figure 4.2.10A. Many peaks that migrated were GB1 peaks, with some examples highlighted by red boxes, suggesting Smad 3 interaction with GB1. However, previous GB1 fusion protein titrations with the same peptide sequence showed minimal interaction of Smad 3 with GB1 (Wahl, 2016). Furthermore, the majority of peak migration is towards the centre of the spectrum. Clustering of peaks in this manner, towards the centre of the spectrum, is characteristic of unfolded proteins due to increased hydrogen exchange with water hence lower ^1H dispersion. Therefore, these results are more indicative of protein denaturation rather than peptide interaction. Further analysis was carried out to determine the cause of this denaturation. Due to the use of synthetic peptides, ^{19}F -NMR was used to find out whether the presence of trifluoroacetic acid (TFA) from the synthetic peptide synthesis process could be the problem. Previous work with synthetic peptides purchased as TFA salts showed no denaturation (Wahl, 2016), however the ^{19}F -NMR for the titration sample suggests that there is a high concentration present in this experiment (Figure 4.2.10B). This explained the denaturation of GB1-WW1 at the later titration points. Samples of peptide used previously may have had a much lower concentration of TFA hence the lack of denaturation. Analysis of the titration points before GB1 peak migration shows residue 87 Asn increasing in intensity suggesting a transition from more than one unbound conformation to a bound conformation with increasing proportions in the bound state resulting in an observed change to the fast exchange regime. 88 Thr also presents a small shift in peak position, and 92 Thr appears to transition from an unbound conformation to a mixed population of bound and unbound presenting as intermediate exchange during the timescale of the experiment (Figure 4.2.11A). These changes are shown in closer detail in Figure 4.2.11B. The changes in 87 Asn and 92 Thr suggests similar interaction residues in WW1 for Smad 2 and Smad 3 peptides, however the lack of movement by 88 Thr suggests either there is a slight difference in binding interface between Smad 2 and 3 peptides with WW1 or that the movement of 88 Thr was not observed for Smad 2 because the titration had not reached the equivalent ratio. The latter is more likely due to only 2 differences in the amino acid sequences for Smad 2 and Smad 3 peptides.

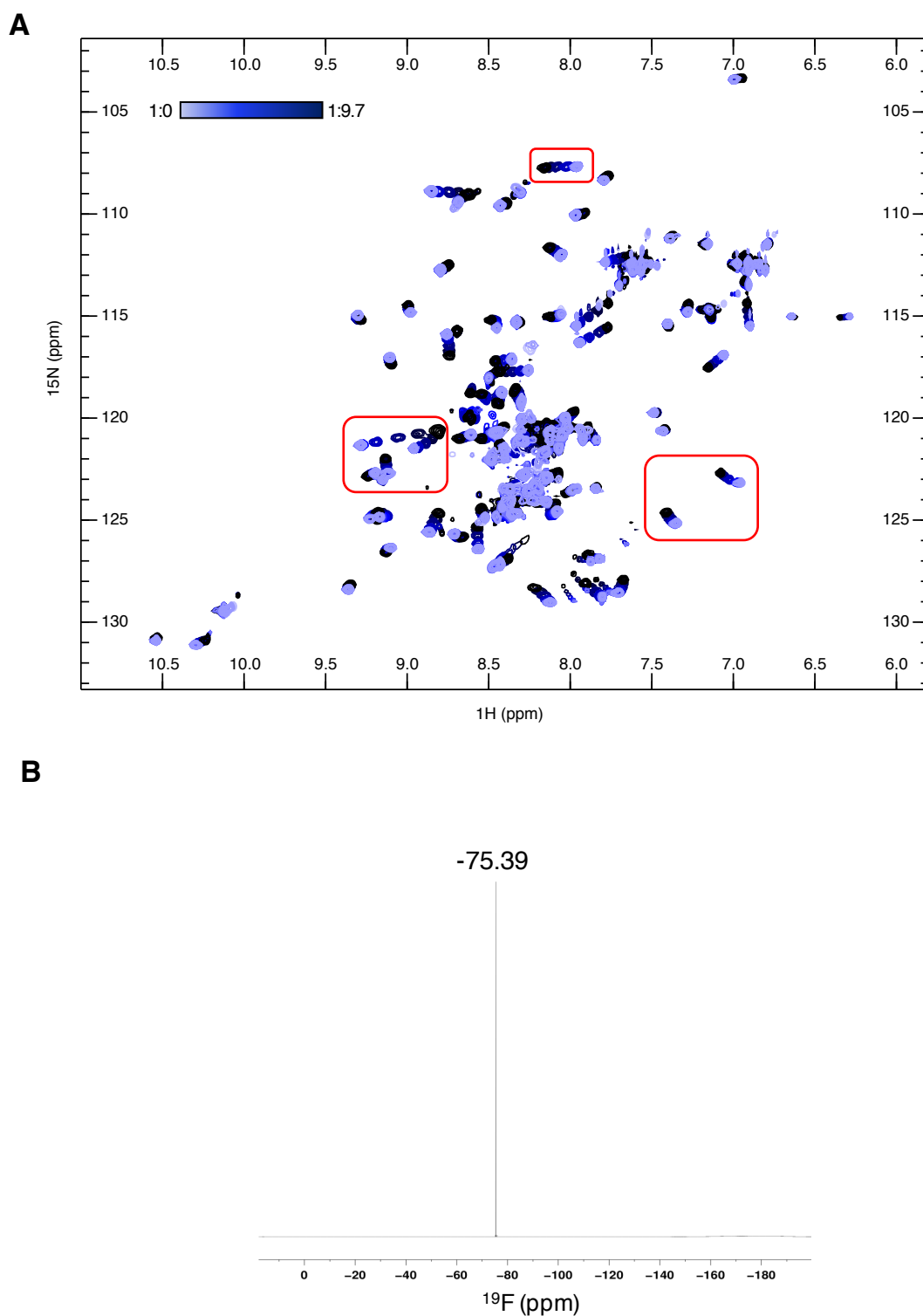


Figure 4.2.10 - ^{15}N -GB1-WW1 titration with increasing Smad 3 peptide additions.

A) Overlaid HSQCs of titration points 1:0 (light blue) to 1:9.7 (black). Red boxes indicate examples of GB1 peak movement. B) ^{19}F -NMR spectrum of ^{15}N -GB1-WW1:Smad3 after the final titration point. The concentration of ^{15}N -GB1-WW1 used in the titration was 0.6 mM, with a final concentration of 5.85 mM Smad 3 peptide. ppm – parts per million.

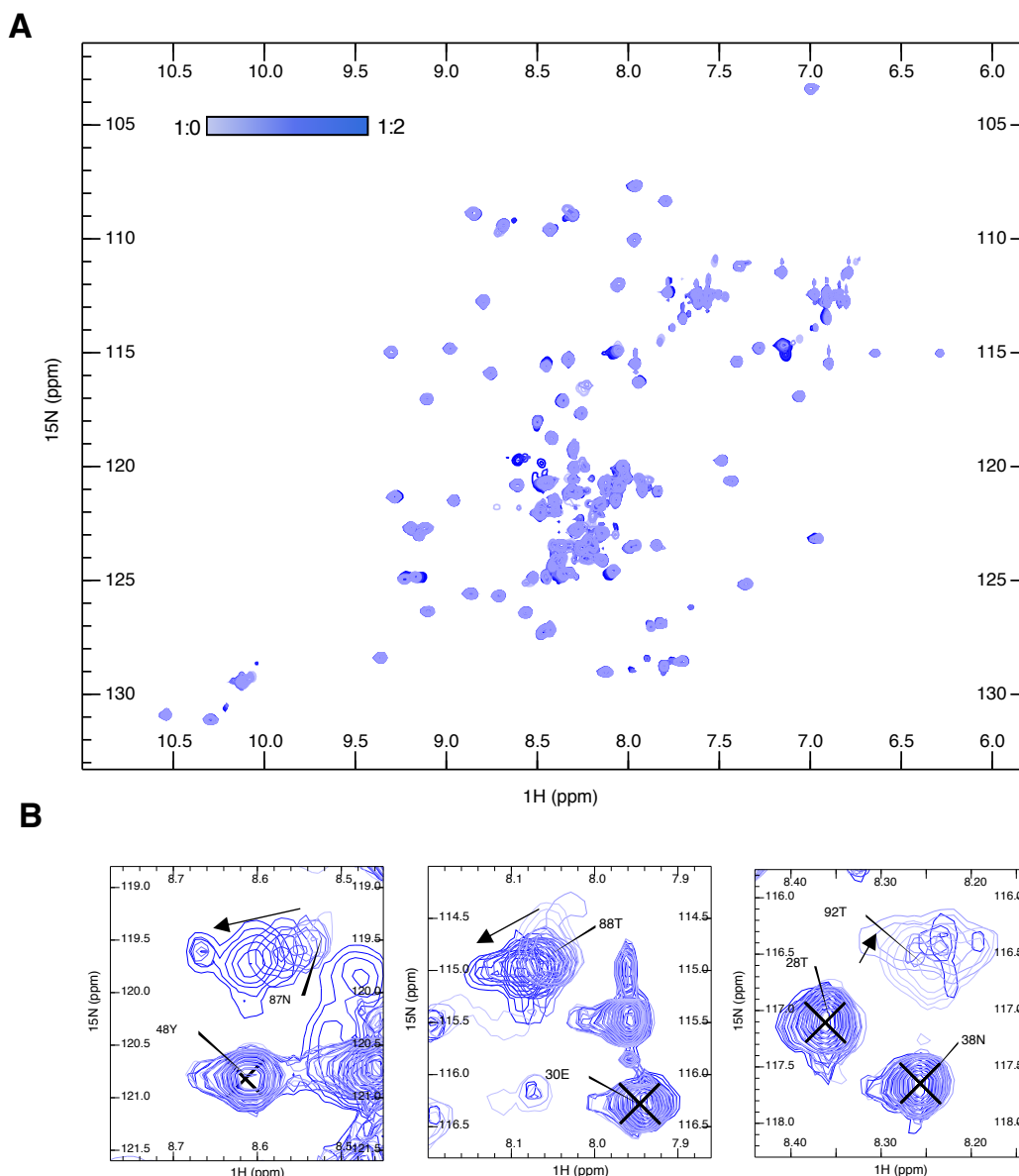


Figure 4.2.11 - Titration points 1:0 to 1:2 for ^{15}N -GB1-WW1: Smad 3.

A) ^{15}N -GB1-WW1 titration with synthetic Smad 3 peptide, ranging from 1:0 (light blue) to 1:2 (dark blue). B) Spectral regions of amino acids 87 Asn (87N), 88 Thr (88T) and 92 Thr (92T), highlighting the changes observed over the first half of the titration. Arrows indicate direction of migration and X marking no change in position. ppm – parts per million.

4.2.4.3. GB1-WW1 titration with Smad 7 peptide

Lastly, Smad 7 peptide was used to confirm the lack of binding observed for WWP2-N by investigating the interaction between WW1 and Smad 7 peptide. Due to previous titration analysis of WW4 with Smad peptides showing interaction capabilities with Smad 2 and 3 as well as 7, but at a lower affinity for the R-Smads (Wahl, 2016), the analysis here was used to determine whether an inverse interaction pattern to WW4,

where lower affinity is present for the I-Smad instead, could exist between WW1 and Smad 7. A titration was set up as described in section 2.4.3 and HSQC spectra acquired at 800 MHz. Comparison of all titration points showed the appearance of a few sharp peaks around $^{15}\text{N} \sim 130$ ppm, $^1\text{H} \sim 8$ ppm (Figure 4.2.12), suggesting the presence of degradation. Due to previous data showing no interaction of Smad 7 with WWP2-N and the location of these peaks towards the bottom of the spectra, it is more likely that these peaks are due to degradation rather than folding of WW1 upon Smad 7 binding. However, the appearance of peaks in the central cluster and beside 70 Gly and 79 Gly at $^{15}\text{N} \sim 110$ ppm with 79 Gly disappearing from the later spectra, and the appearance of 88 Thr in later HSQCs suggests that there is a change in folding. It is difficult however to conclude that there could be weak interaction of Smad 7 with WW1, because of the lack of significant changes, and presence of degradation.

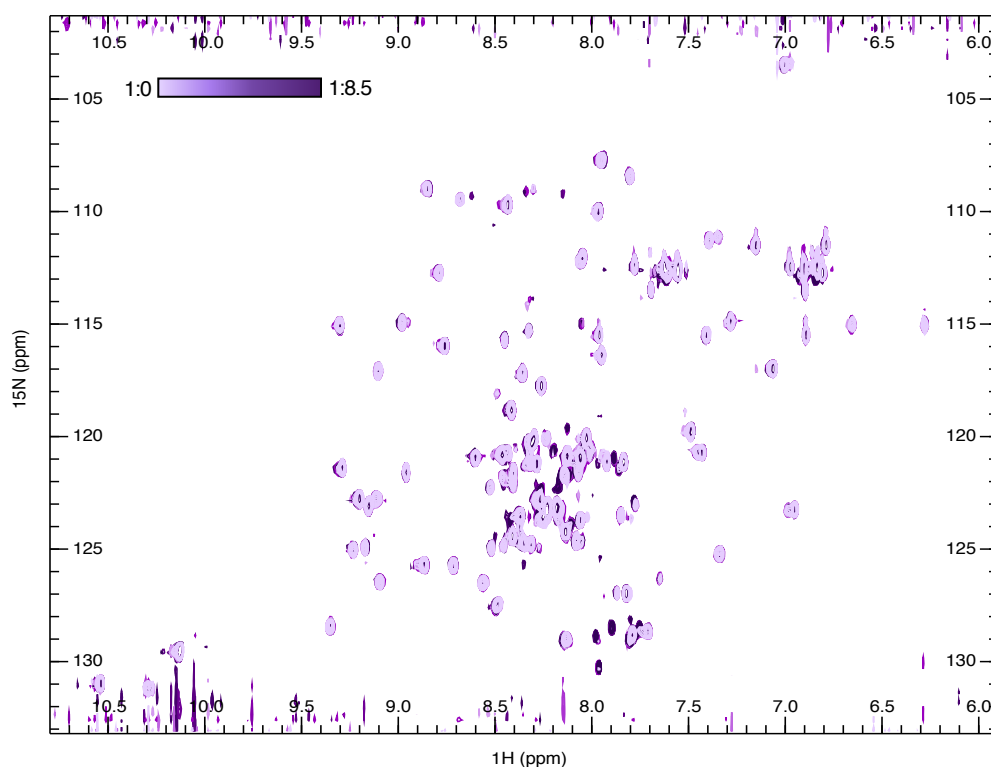


Figure 4.2.12 - Titration HSQCs of ^{15}N -GB1-WW1 with increasing concentrations of Smad 7 peptide.

Overlaid HSQC spectra for Smad 7 peptide titration points ranging from 1:0 (light purple) to 1:8.5 (dark purple). 0.06 mM ^{15}N -GB1-WW1 was used throughout the titration with a final concentration of 0.51 mM Smad 7 peptide at 1:8.5 ratio. ppm – parts per million.

4.2.5. Partial assignment of ^{15}N -GB1-WW1-2

Studies have suggested coupled activity of WW domains in other Nedd4 E3 ligases, which had not yet been analysed in WWP2. The end of WW1 and start of WW2 overlap in sequence slightly and therefore lack an obvious linker region between the two domains, suggesting that perhaps when both are present within WWP2-FL they may function in a cooperative manner unlike when WW1 is isolated in the WWP2-N isoform. In order to begin investigations into whether WW domains 1 and 2 influence each other when present in WWP2-FL, GB1-WW1-2 fusion protein was bacterially expressed and purified as described in section 2. $[^1\text{H}\text{-}^{15}\text{N}]$ -HSQC analysis to look at the amide group positions was carried out to determine if there were any significant differences compared to GB1-WW1 and GB1-WW2. GB1-WW1-2 has a total of 124 assignable residues excluding prolines, however only 98 peaks excluding side chains were present in the HSQC (Figure 4.2.13). This suggests the other 34 peaks are in intermediate exchange and unobservable under the conditions and timescale of the experiment. As there were too many peaks in intermediate exchange to acquire triple resonance data on GB1-WW1-2, the partial assignments for WW1 (Figure 4.2.7) and WW2, provided by Danielle Bourcier, were used to try and assign some of the peaks in WW1-2. Only 75.8 % of peaks were assignable, with the majority being GB1 peaks. When the HSQC spectrum for ^{15}N -GB1-WW1-2 was compared with ^{15}N -GB1-WW1 and ^{15}N -GB1-WW2, there were several differences observed (Figure 4.2.14). The majority of GB1 peaks were identical between the constructs, except for some deviation seen for residues 6 Tyr and 23 Ala in GB1-WW2. Several peaks showed no overlap with those from GB1-WW1 or GB1-WW2 suggesting WW1-2 has a different fold to either independent domain. One similarity across all three constructs though was missing peaks and intermediate exchange. This may suggest that these domains have an intrinsically unfolded structure when unbound.

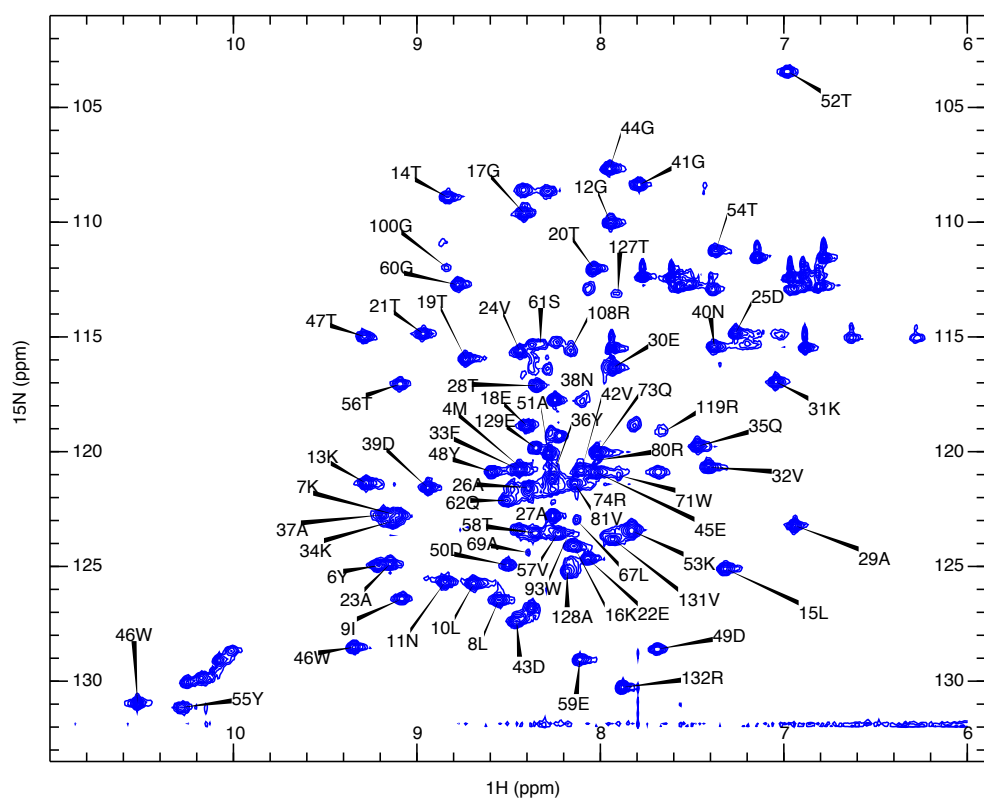


Figure 4.2.13 - $[^1\text{H}-^{15}\text{N}]$ -HSQC with partial backbone assignment for ^{15}N -GB1-WW1-2.

Annotated HSQC of GB1-tagged WWP2 WW1-2 tandem domain with the backbone amide group assignment where possible. ppm – parts per million.

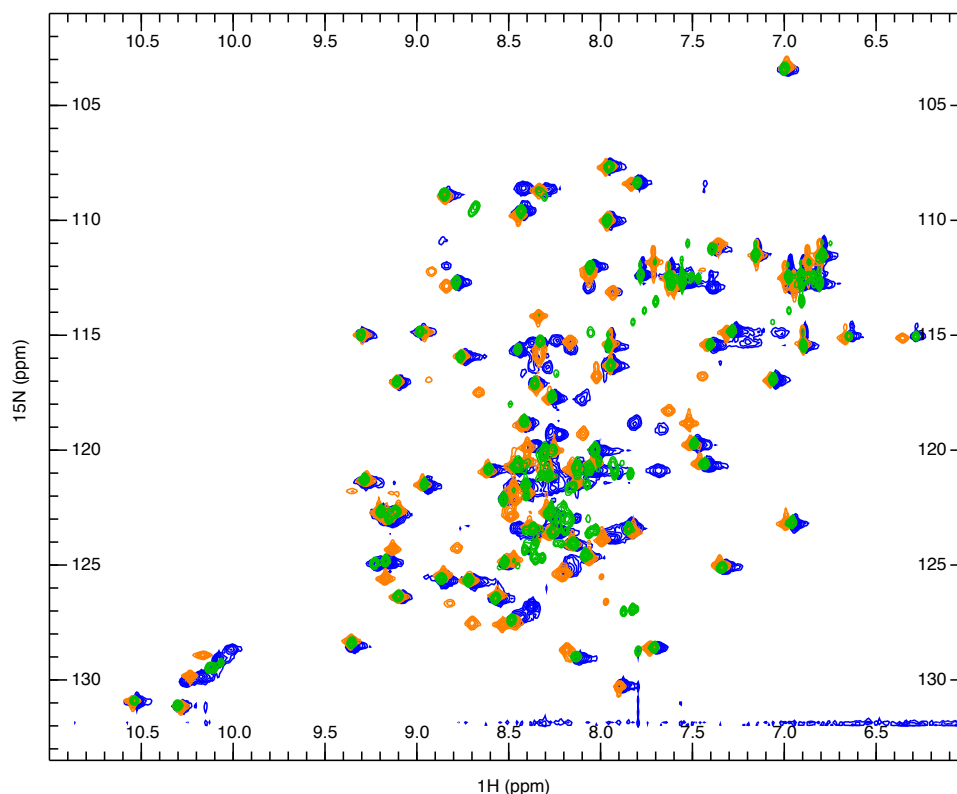


Figure 4.2.14 - GB1-WW1, -WW2 and -WW1-2 HSQC spectra overlaid.

Single-labelled GB1-tagged WW1 (green), WW2 (orange) and WW1-2 (blue) HSQC spectra. ppm – parts per million.

4.2.6. Titration analysis of ^{15}N -GB1-WW1-2 tandem domain with Smad-based peptides

The amino acid sequence for WW1-2 has often been depicted by having a linker region between the domains. Closer analysis of the end point of the N isoform and the start of the WW2 domain however suggests that there could potentially be an overlap of the 330-RPLP-333 sequence at the end of WW1 and start of WW2. This suggests that these domains may influence each other in FL isoform, unlike when WW1 is isolated in the N isoform, possibly resulting in different outcomes. The next section looks into the binding of Smad peptides to the tandem domain WW1-2, with comparison to WW1 and WW2 individually.

4.2.6.1. GB1-WW1-2 titration with Smad 2 peptide

To look at the binding of Smad 2 peptide with WW1-2, ^{15}N -GB1-WW1-2 was bacterially expressed and purified as described in sections 2.2.2 and 2.2.4. Sequential additions of peptide were made to the sample and $[^1\text{H}-^{15}\text{N}]$ -HSQCs acquired after each addition, up to a ratio of 1:10 protein to peptide. Comparison of the HSQCs showed the same

pattern of GB1 migration (examples highlighted in red boxes in Figure 4.2.15A) as observed for GB1-WW1 with Smad 3 peptide previously discussed in section 4.2.4.2. ^{19}F -NMR confirmed the presence of TFA in this sample of peptide (Figure 4.2.15B), resulting in denaturation of ^{15}N -GB1-WW1-2 after a 1:4 ratio of protein to peptide. Analysis of the titration before denaturation though, showed some very interesting changes, highlighted in Figure 4.2.16A. One significant aspect is the migration of the tryptophan side chains located to the bottom left region of the spectra (Figure 4.2.16A and B), suggesting the side chains are changing chemical environment, perhaps due to peptide binding inducing protein folding or the result of direct binding as the expected binding site, the XP pocket, contains a tryptophan residue. It is clear that 2 side chains undergo a greater perturbation in chemical shift than the other 2. As there are currently no side-chain assignments for WW1-2, using the overlaid spectra of WW1 and WW2, it is clear that one peak which showed a reduced degree of migration was located in WW2, however it is unclear which of the other 3 belong to each domain. The movement of the tryptophan side-chains was not present in the previous WW1 titrations suggesting that the presence of WW2 has either altered the residues involved in binding or an increase in the affinity of the domains for Smad 2 peptide has occurred. However, elucidating which is the case requires further assignments and experimentation. Another observation was the appearance of new peaks, as shown by the purple boxes in Figure 4.2.16A. It was clear that residues 100 Gly, 108 Arg and 119 Arg were shifting as the peptide concentration increased (Figure 4.2.16C showing 100 Gly and 119 Arg in closer detail). This suggests WW2 within the WW1-2 construct is capable of binding to Smad 2 peptide, which was previously unknown. Intriguingly, no significant changes were observed for 87 Asn and 92 Thr which were missing throughout the titration or present at an alternative position, further supporting a difference between the interaction and folding of WW1 only and WW1-2 in the presence of Smad 2 peptide. These domains may be worth further investigation by collecting backbone assignments, which was not carried out here due to time restrictions, for the bound tandem domain as there is an increased number of peaks to 114 by ratio point 1:4 and in turn could possibly increase further if the protein becomes saturated with peptide, potentially inducing complete folding.

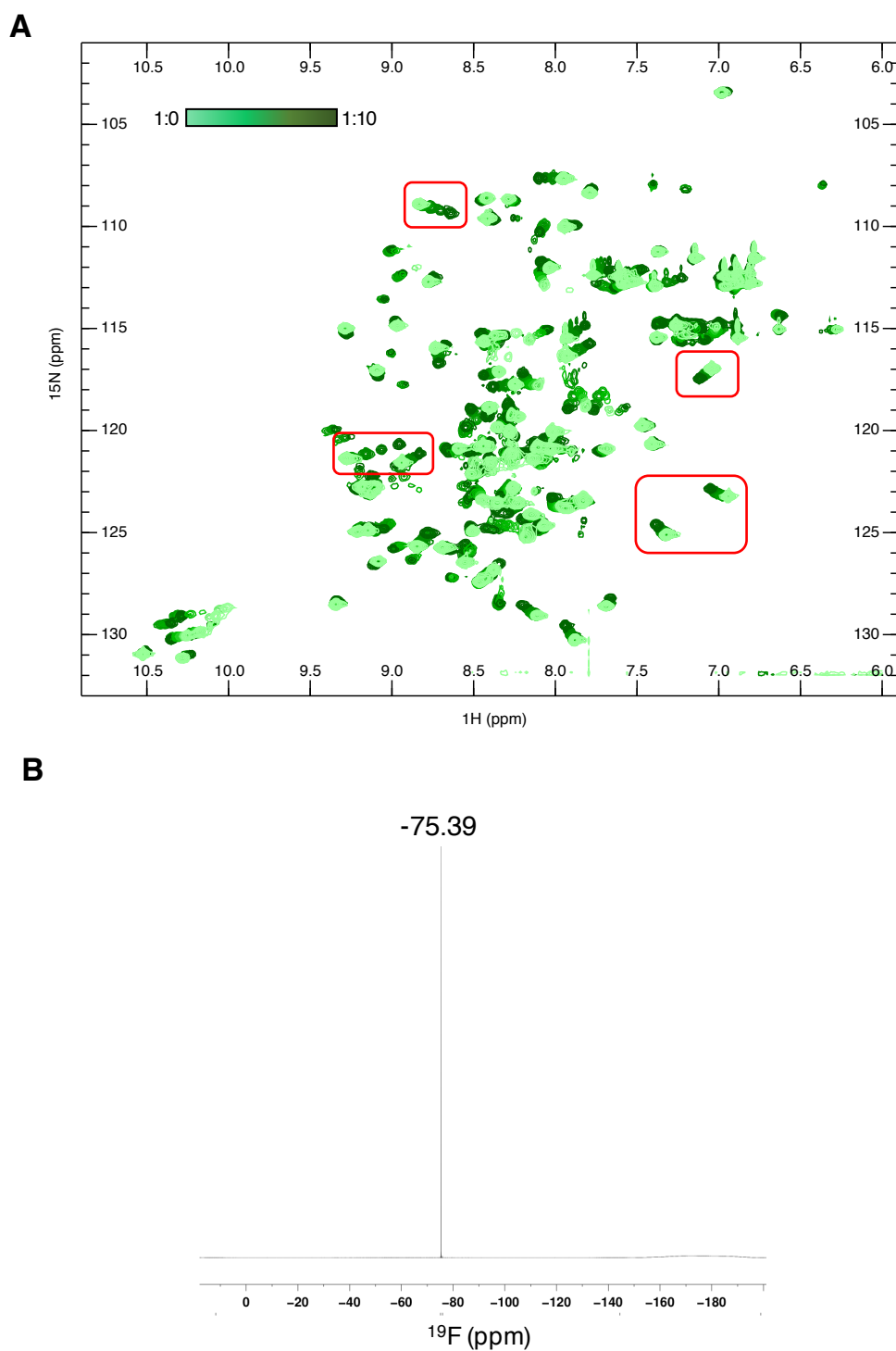


Figure 4.2.15 - Titration analysis of ^{15}N -GB1-WW1-2 with Smad 2 peptide shows protein denaturation.

A) $[\text{}^1\text{H}-^{15}\text{N}]$ -HSQC spectra overlaid for titration points 1:0 (light green) to 1:10 (dark green). Red boxes show examples of GB1 peak perturbations. B) ^{19}F -NMR of the sample used in A. ^{15}N -GB1-WW1-2 was 0.72 mM across titration points, with a final Smad 2 peptide concentration of 7.2 mM. ppm – parts per million.

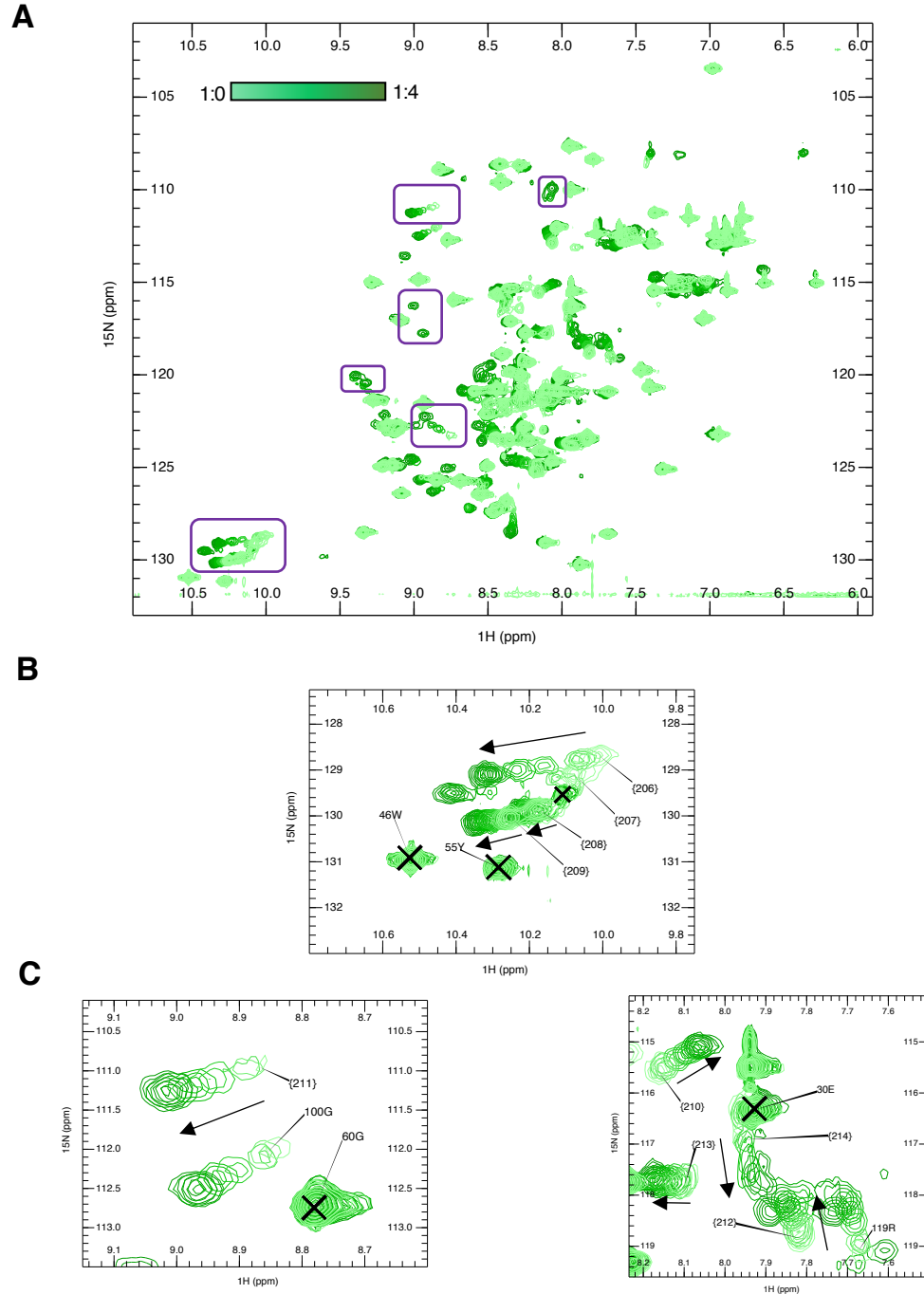


Figure 4.2.16 - $[^1\text{H}-^{15}\text{N}]$ -HSQCs from 1:0 to 1:4 for ^{15}N -GB1-WW1-2 titration with Smad 2 peptide before denaturation.

A) Titration points 1:0 (light green) to 1:4 (dark green) for ^{15}N -GB1-WW1-2 with increasing concentrations of Smad 2 peptide. Purple boxes highlighting new peaks, and tryptophan side chain migrations. B) Spectral region of the tryptophan side-chains showing non-moving GB1 tryptophan side chain and 55 Tyr peaks compared to WW1 and WW2 tryptophan side-chains perturbation. C) 100 Gly and 119 Arg regions showing peak shifts and new peak appearance. Arrows indicate direction of movement, and X marks non-moving peaks. ppm – parts per million.

4.2.6.2. ¹⁵N-GB1-WW1-2 titration analysis with Smad 3 peptide

Similar to previous experiments, ¹⁵N-labelled GB1-WW1-2 was expressed and purified as outlined in sections 2.2.2 and 2.2.4. Smad 3 peptide was added to the NMR sample tube in amounts equivalent to a ratio of 1:0 up to 1:9.8 of protein to peptide (Table 2.4.9). [¹H-¹⁵N]-HSCQs were acquired at every titration point. Unlike the ¹⁵N-GB1-WW1 titration with Smad 3 peptide, which used a different synthetic peptide sample, and the Smad 2 titration with ¹⁵N-GB1-W1-2, there was very little movement of GB1 which was only observed for the last titration point (Figure 4.2.17), and therefore this point was not included in further analysis. Several changes in peak positions for residues within WW1-2 were observed as the peptide ratio increased, as seen in Figure 4.2.18A. In a similar manner to the Smad 2 titration (Figure 4.2.16), residues 100 Gly, 108 Arg and 119 Arg showed a transition from unbound conformations in intermediate exchange towards a bound conformation, suggesting that these residues in WW2 are involved in R-Smad peptide binding. The identification of the same residues was not surprising due to the amino acid sequences for Smad 2 and 3 peptides being very similar. The migration of tryptophan side-chains in an identical pattern to the Smad 2 titration further supports these residues as general R-Smad interaction points. In contrast, several peaks appeared around 31 Lys (¹H~7ppm, ¹⁵N~117ppm). This region is shown in closer detail in Figure 4.2.18B overlaid with the same region for the WW1-2: Smad2 peptide titration at approximately equivalent ratios. The lack of overlap for the majority of these peaks between the two titrations initially suggested that these residue peaks may be an important difference between the Smad 2 and 3 interactions, however further analysis showed the peaks which failed to overlap in the original spectrum were aliased peaks which overlap at ~85 ppm (Figure 4.2.18B, right panel), characteristic of arginine NεHε side chain peaks. As arginine side chain amide peaks are only observed under certain conditions, this suggests that either the pH of the samples became more acidic over the course of the titration, which is unlikely as no water peak movement was present for the Smad 3 titration. Another possibility is that these arginine side-chains may be involved in hydrogen-bonding in the presence of peptide or they are being buried by the folding of the domain. The new peaks which overlap in the original spectrum are most likely main-chain NH peaks as they are not pairs, which then suggests that these residues are potentially involved in another folding arrangement induced by the specific Smad that is bound. These possibilities require further investigation as understanding the differences may reveal how the selectivity of WWP2-FL towards particular Smad proteins is controlled.

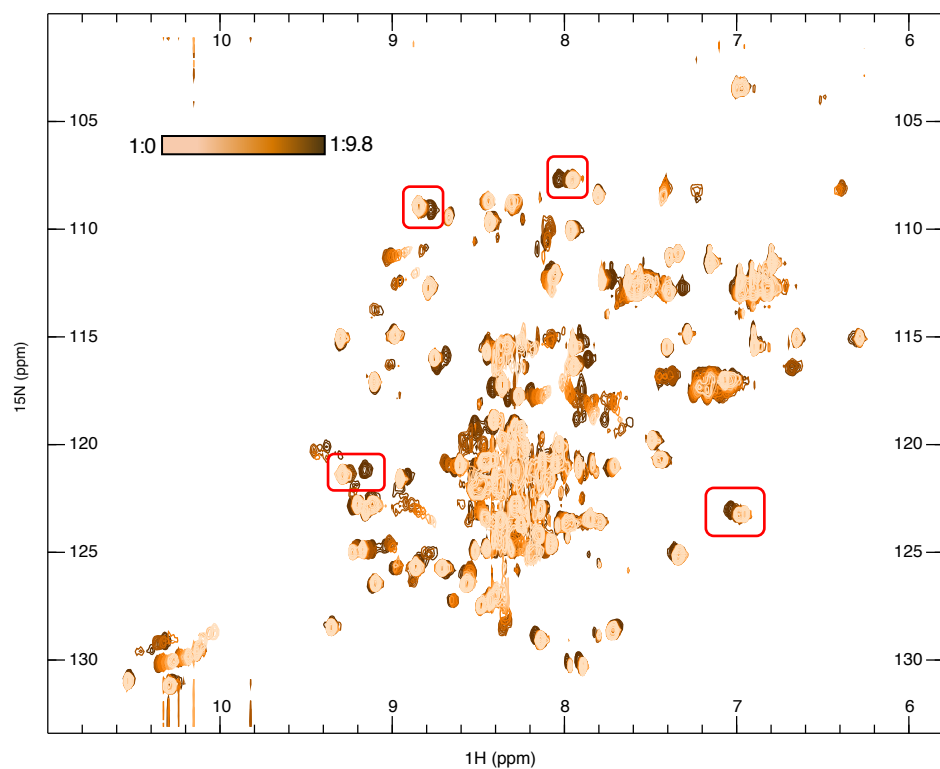


Figure 4.2.17 - ^{15}N -GB1-WW1-2 titration with increasing Smad 3 peptide concentration leads to changes in peak chemical shifts.

Overlaid $[\text{}^1\text{H}\text{-}^{15}\text{N}]$ -HSQCs starting with ratio 1:0 of ^{15}N -GB1-WW1-2 to Smad 3 peptide up to a ratio of 1:9.8. Minimal GB1 peak movement highlighted by red boxes. 0.3 mM ^{15}N -GB1-WW1-2 was used and a final Smad 3 peptide concentration of 2.95 mM. ppm – parts per million.

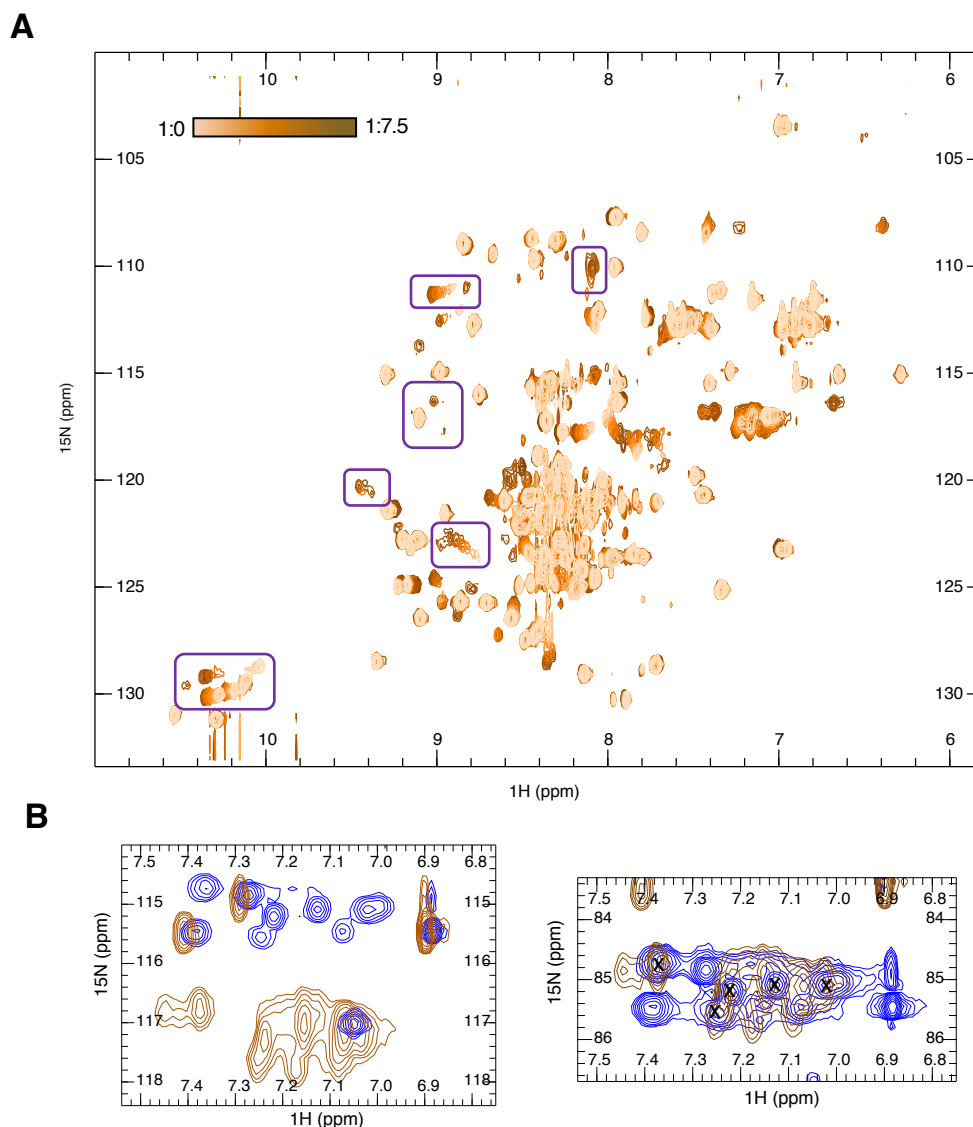


Figure 4.2.18 - Increasing Smad 3 peptide concentration leads to chemical shift perturbation of ^{15}N -GB1-WW1-2.

A) $[^1\text{H}-^{15}\text{N}]$ -HSQC spectra overlaid from 1:0 to 1:7.5 of WW1-2: Smad 3 peptide, excluding the final point due to GB1 movement. Purple boxes highlighting peak appearance as seen in Smad 2 titration as well. B) Spectral regions showing aliased peaks present at ratio 1:4 in the Smad 2 (blue) titration and ratio 1:7.5 for the Smad 3 (brown) titration. X marks the folded peaks.ppm – parts per million.

4.2.6.3. GB1-WW1-2 titration analysis using Smad 7 peptide

The final titration with Smad 7 peptide was carried out to investigate whether the presence of WW2 may alter the role of WW1, and vice versa, in relation to Smad 7 binding. Previous work has shown that Smad 7 binding to WW1 was not clearly observed in the titration in section 4.2.4.3 but Smad 7 has been shown to interact with WW2 and induce folding (Danielle Bourcier, personal communications). ^{15}N -labelled

GB1-WW1-2 was prepared as described in sections 2.2.2 and 2.2.4. No movement of GB1 was detected throughout this titration meaning all changes observed were due to the binding of Smad 7 peptide (Figure 4.2.19A). Several new peaks were present at the end of the titration suggesting folding of the tandem domains into a bound conformation upon Smad 7 binding. Interestingly, the tryptophan side chain peaks and a few other peaks (Figure 4.2.19B) in this titration disappear and re-appear, indicative of a transition from a single unbound conformation to a bound conformation, rather than having a traceable migration path as observed in the Smad 2 titration. There were 124 peaks present for ^{15}N -GB1-WW1-2: Smad 7 at the last titration point. When taking proline residues into account, this indicates that the amide groups were all visible by the end of the titration in the experimental timeframe, suggesting that the tandem domain underwent complete folding in the presence of Smad 7. The appearance of peaks in similar positions to those present in the Smad 7 bound state of GB1-WW2 (examples highlighted by blue boxes in Figure 4.2.19A) (GB1-WW2: Smad 7 work carried out by Danielle Bourcier), suggests that even in the presence of WW1, the chemical environments of most residue amide groups in WW2 are not dissimilar to that of WW2 only in presence of Smad 7. Furthermore, the migration of all four tryptophan side chains supports the folding of WW1-2 in the presence of Smad 7 peptide as well as direct peptide binding, due to the 1st tryptophan residue of each domain being associated with domain folding and stability, and peptide binding involving the 2nd tryptophan residue of each domain residing in the XP pocket.

Comparison of all three GB1-WW1-2 titrations highlighted peaks unlikely to converge upon saturation, such as the peaks on the left side of the spectra (^1H - ~ 9.5 , ^{15}N - ~ 120 -125) (Figure 4.2.20A, purple box) and two of the four tryptophan side chain amide groups (Figure 4.2.20A, blue box). This shows that slightly different chemical environments were observed between the R-Smads and I-Smad, supporting the possibility that when both domains are present in WWP2-FL it may be the concerted actions of WW1 and WW2 that determine which type of Smad protein is targeted and when, especially since WW1 appears to respond differently as a tandem domain with Smad 7 compared to on its own. As for the aliased peaks, only 6 out of 8 possible arginine residue peaks are present for WW1-2. These peaks are observed in all 3 titrations with GB1-WW1-2 (Figure 4.2.20 A and B). When compared to the GB1-WW1 titrations, only a single peak in this region appeared in the Smad 2 and Smad 3 titrations with WW1, but not in the Smad 7 titration which used the same peptide sample. As the side chain amide peaks broaden in the pH range of pH 6 – 7.2 (Yavari and Roberts,

1978), and the pH of the GB1-WW1-2: Smad 7 titration changed slightly by -0.25 from an initial pH of 6.5, this suggests that even with the decrease in pH it is unlikely that these peaks should be present. Hence it is possible these are the consequence of hydrogen bonding. Furthermore, these peaks are not present in the Smad 7 titration with GB1-WW2 (Danielle Bourcier, personal communication), suggesting that this may be a tandem domain result with regards to the presence of WW1. It is, however, important to remember that as saturation was not reached with Smad 2 and Smad 3 peptides as previously mentioned, the comparison between the 3 substrate peptides with GB1-WW1-2 should be taken lightly, and further analysis and data is necessary to draw more accurate conclusions.

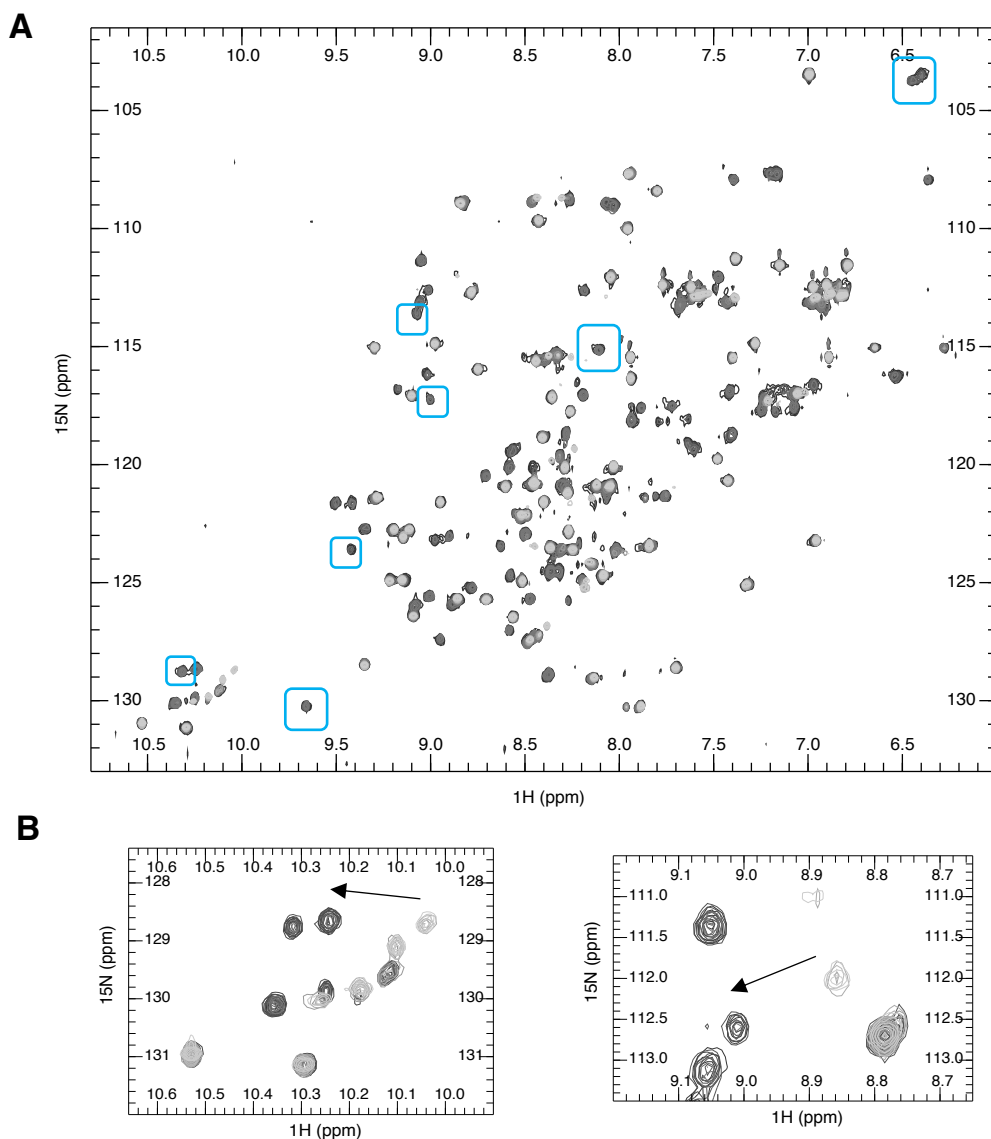


Figure 4.2.19 - $[^1\text{H}\text{-}^{15}\text{N}]$ -HSQC spectra of increasing Smad 7 peptide concentration against ^{15}N -GB1-WW1-2.

A) Ratio 1:0 (light gray) to 1:11 (black) and intermediate ratio $[^1\text{H}\text{-}^{15}\text{N}]$ -HSQCs for ^{15}N -GB1-WW1-2 with the addition of Smad 7 peptide at each point. The final concentration of Smad 7 peptide used was 1.7 mM, with 0.15 mM ^{15}N -GB1-WW1-2 used in the titration. Blue boxes highlight example WW2 peaks that are identical in chemical shifts between the WW2 and WW1-2 titrations with Smad 7 peptide. B) Examples of peaks which disappeared and reappeared across the titration points – left panel shows the tryptophan side chain region; right panel shows examples of other peaks. ppm – parts per million.

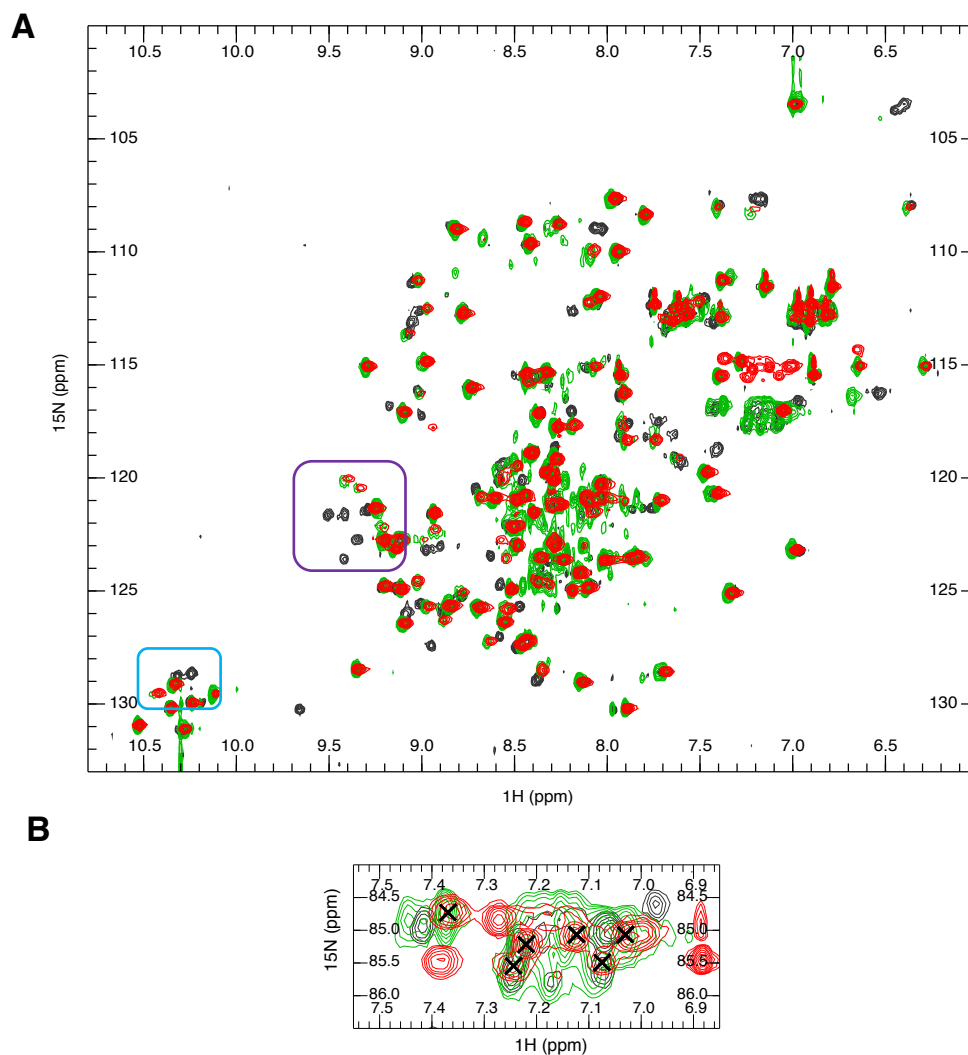


Figure 4.2.20 - Comparison of $[^1\text{H}-^{15}\text{N}]$ -HSQC spectra of WW1-2: Smad7 at 1:11 with WW1-2: Smad 2 1:4 and WW1-2: Smad 3 1:7.5.

A) Overlaid spectra for GB1-WW-1-2 titrations with Smad 2 (red), 3 (green) and 7 (black) at the points just before GB1 movement or final titration point. Purple boxes show an example of different peak positions between the spectra. B) Comparison of the spectral region around ~85 ppm for the same spectra as part A, where X denotes the aliased peaks. ppm – parts per million.

4.3. Discussion

Understanding the relationship between structure, binding and function is vital to the development of effective therapies. Many pharmaceutical companies utilise information regarding selective protein-protein interactions to direct the design of peptide-based drugs that mimic biologically relevant interactions associated with disease. WWP2 resides in a group of HECT E3 ligases associated with oncogenesis when overexpressed. The domain architecture of WWP2 makes it ideal for novel therapeutics with its four WW domains that selectively target tumour suppressor proteins, as described previously and in more detail in chapter 1. The aim of this chapter was to utilise pull-down and NMR titration analysis to understand WW domain interactions with specific substrates in the hopes of enabling the design of potential inhibitory peptides.

The high prevalence of PTEN loss in prostate cancer by both genetic abnormalities and protein-associated depletion indicates the importance of PTEN tumour suppressive activities (Wang *et al.*, 1998). Previous studies by Trotman *et al.* (2007) and Wang *et al.* (2007) suggested that Nedd4-1 was the main E3 ligase for PTEN. Evidence suggested that Nedd4-1 controlled both the nuclear shuttling of PTEN via mono-ubiquitination and the labelling of PTEN for degradation in the cytoplasm via poly-ubiquitination. The regulation of cellular localisation and turnover by Nedd4-1 instils E3 ligase activity as a switch for the regulation of PTEN tumour suppressor activity shown in studies involving prostate cancer cell lines as well as mouse model xenografts (Wang *et al.*, 2007). The increase in PTEN accumulation associated with Nedd4-1 mediated ubiquitination of PTEN resulted in suppression of AKT phosphorylation and in turn downstream signalling (Trotman *et al.*, 2007). This suggested that the recovery of PTEN levels in cancer cells by preventing ubiquitination-mediated degradation may be a possible method of therapy. Later work published by Maddika *et al.* (2011) established a second HECT E3 ligase, WWP2, as a regulator of PTEN activity via polyubiquitination. The knockdown of WWP2 was also shown to increase the endogenous levels of PTEN and in turn decrease phospho-AKT and prostate cancer cell line survival. In contrast, overexpression led to decreased PTEN and increased proliferation. At the time of this research by Maddika and colleagues was the first confirmation of WWP2 isoform activity in cancer development. Evidence of a role of other WWP2 isoforms in the regulation remains unanswered despite the data supporting WWP2 as the predominant regulator of PTEN activity. Initial investigations in this project attempted to uncover the sites of PTEN interaction on WWP2 in order to isolate a specific region of WWP2 for use in further studies. Pull-down

analysis of PTEN and WWP2 constructs revealed the potential for a wider interaction site across both proteins. WWP2-C showed binding to the C2 domain of PTEN, whereas WWP2-N bound to the phosphatase domain. This suggested that the binding of isoforms C and N to PTEN may indicate a role for these isoforms in the control of PTEN, adding to the already complex regulatory network. The binding of the N isoform to the phosphatase domain is in parallel with WWP2-FL as shown by Maddika *et al.* (2011). This suggests that the N isoform may function to prevent the interaction of WWP2-FL with PTEN. Alternatively, WWP2-N may actually recruit PTEN moieties to the full-length protein increasing the likelihood of PTEN ubiquitination. Georgescu *et al.* (1999) analysed the degradation rates of PTEN and PTEN mutants and found that while there was a constant decrease in the degradation of mutants, the levels of PTEN-WT stabilised. They proposed that sequestration of new PTEN moieties could be preventing their degradation. It would be interesting to determine whether WWP2-N is involved in this observation or whether some other factor is involved in PTEN steady-state expression levels. Further work is therefore required to determine the biological function behind these interactions.

In contrast, binding of WWP2-C to C-terminal tail of PTEN was observed. Interestingly, this region of PTEN conveys stability and acts as an inhibitor of PTEN activity (Odriezola *et al.*, 2007; Vazquez *et al.*, 2000). Furthermore, the tail of PTEN contains a cluster of serine and threonine residues that undergo phosphorylation important for function. In conjunction with the loss of the tail promoting PTEN activity yet reducing half-life, mutation of these phosphorylation sites also resulted in a decrease in stability and half-life, and an increase in activity (Vazquez *et al.*, 2000). This suggests that this region of PTEN is important in the regulation of degradation. Interaction of WWP2-C with the C-terminal tail of PTEN may be the interaction promoting the rapid degradation of PTEN when the tail is unphosphorylated or mutated. Whether this interaction occurs with the phosphorylated form of PTEN has yet to be established. However, Chen *et al.* (2016) observed a marked decrease in WWP2-FL-mediated ubiquitination when the C-terminal tail of PTEN was phosphorylated. This opens up more questions as to the role of WWP2 isoforms in PTEN regulation and the difference in binding sites observed may confer different consequences based on the cellular environment and PTEN status.

Further investigation into WWP2 interactions with PTEN used pull-down analysis with tandem WW domains 1-2 and 3-4 to determine whether these domains were involved

in WWP2-FL interaction. Jiang *et al.* (2015) stated that the WW domains of WWP2 were exclusively PPxY motif binding domains, however binding was observed for both tandem domains to PTEN constructs. WW1-2 bound to the C2 domain or C-terminal tail, whereas WW3-4 seemed to interact between amino acids 1-100. One possibility for the opposing sites of interaction observed for WW1-2 and WW3-4 compared to the isoforms could be that the WW domains 1 and 4 in the N and C isoforms, respectively, may bridge across to interact with the other domain of PTEN, in turn stabilising the interaction. The overall size of PTEN at 430 amino acids, and the fairly compact structure solved by X-ray crystallography (PDB: 1D5R, Lee *et al.*, 1999; lacks the first 13 residues and the C-terminal tail) suggests the possibility of interaction with both domains. Although evidence from interaction analysis of the yeast Nedd4-like ubiquitin ligase Rsp5 has implied the HECT domain could be involved in substrate interactions (Lee *et al.*, 2009) this has yet to be shown for human Nedd4 E3 ligases, but the results of this pull-down analysis and lack of WW domain binding motif in PTEN may turn out to be a cooperative binding event of multiple domains within WWP2. The other possibility that cannot be excluded is an independent function of these WW domains as tandem domains. Emerging evidence of the concerted action of WW domains for other Nedd4 E3 ligases (Aragón *et al.*, 2012; Chong *et al.*, 2010) and preliminary data for WW3-4 of WWP2 (Wahl, 2016) suggests these domains may influence the binding events of each domain when present together. However, as this work only looked at whether there was binding or not and not function or interaction kinetics with PTEN, this requires further exploration. Lastly, one aspect that cannot be dismissed is that the tandem WW domains and isoform constructs were used as whole cell lysates in the pull-downs, therefore the role of adaptor proteins facilitating the interactions with PTEN should be taken into consideration. For example, NDFIPs are known adaptor proteins for Nedd4 E3 ligases which can either act to relieve auto-inhibition or facilitate the recruitment and interaction between substrates and Nedd4 E3 ligases (Riling *et al.*, 2015; Trimpert *et al.*, 2017). Uncovering the binding sites of WWP2 isoforms to PTEN could reveal a whole new area of interactions and functional roles for WWP2 that can explain why WWP2 ubiquitinates PTEN and in what cellular context.

As discussed above, there appears to more complexity to the PTEN interaction within WWP2 than originally thought. In conjunction with reduced expression levels of truncated PTEN proteins, inclusion bodies of PTEN constructs and lack of an isolated interaction site, it seemed that NMR analysis of this was not achievable at this stage of investigations. Therefore, the focus shifted towards Smad protein substrates due to the

presence of an isolated region of these proteins ideal for NMR analysis. Within the Smad proteins linker region exists a well-known PPxY motif targeted by Nedd4 E3 ligases (Aragón *et al.* 2012; Chong *et al.* 2006; Lin *et al.* 2000; Soond and Chantry, 2001a). While many studies which looked at small regions of interacting partners with WW domains utilised synthetic peptides to understand the structural basis behind the contacts (Aragon *et al.*, 2012; Panwalker *et al.*, 2016), there is evidence that recombinant means can also be used (Chong *et al.*, 2006; Chong *et al.*, 2010). There are advantages to using recombinant peptides which include no limit to the length of peptide, due to the difficulties associated with purification of longer peptides from chemical synthesis, and the ability to isotopically label them for use in structure-based experiments. In this project an attempt to generate similar peptides using recombinant means was analysed, however was unsuccessful. Previous work showed that Smad 7 peptide could be generated with the use of a His-SUMO-tagged construct (Wahl, 2016). This enabled nickel affinity purification, and also utilised the SUMO-tag associated advantage of retaining the true N-terminal region of the peptide, conveyed by the ability of SUMO-protease to cleave the di-Gly with all amino acids except proline at the +1 position (Malakov *et al.*, 2004). Four extra amino acids were included at the beginning for the Smad 7 peptide to increase the molecular weight thus reducing the potential for peptide loss upon dialysis. Thus, N-terminal extensions were used for the Smad 2 and Smad 3 recombinant peptides to account for this aspect of the purification procedure. Malakov *et al.* (2004) established that Ulp-1 tended to work optimally at between 22 and 37 °C and pHs 8-10. As these experiments were run at room temperature and pH 8, it is unlikely that these factors were responsible for the lack of digestion or apparent activity of the enzyme. This therefore suggests that the construct is either too unstable in those conditions or another factor is involved. It is possible, as briefly mentioned previously, that the N-terminal region of the Smad 2 and 3 peptides may be sterically hindering the SUMO protease from gaining access to the di-Gly motif. Future work into this may want to test the presence of a weak denaturant to determine whether this could relieve any tight conformation that could be blocking the digestion site. As these peptides are derived from the linker regions of Smad 2 and 3, this implies no significant secondary structure formation and a flexible nature which would be unlikely to be affected by denaturing conditions, which would argue that steric hindrance may not be the problem, but this is still a set of conditions worth trying at this stage. Rather than focus on developing this and given the increasing literature showing successful use of synthetic peptides, Smad 2 and 3 peptides were purchased instead for NMR-based investigations of WW domain interactions.

No evidence of direct binding between the first WW domain of WWP2 to Smad proteins has been published, although this has been implied from WWP2-N interaction analysis (Soond and Chantry, 2011a). Smads 2 and 3 were identified to interact with WWP2-N, but not Smad 7. This contrasted greatly from the C isoform in immunoprecipitation experiments which only bound to Smad 7 and not Smads 2 and 3. Intriguingly, NMR-based investigation of the isolated WW4 domain with peptides based on the PPxY region of Smad proteins 2, 3, and 7 revealed that WW4 does in fact bind to Smad 2 and 3 but with lower affinities than Smad 7 or phospho-Smad 7 (Wahl, 2016). The biological relevance of these interactions however remained unknown. Based on these investigations, it was thought that perhaps a similar effect may be seen for WW1. Investigations of WW1 also had the advantage of potentially identifying the amino acids involved in interactions for manipulation in peptide design since WWP2-N has previously been identified as a significant contributor to prostate cancer development (Soond *et al.*, 2013).

Initial observations characterised WW1 as one of the less stable WW domains due to the short half-life and incomplete folding shown by missing peaks and the appearance of degradation after 3 days, unlike WW2 and WW4 samples (used by Lloyd Wahl and Danielle Bourcier). Jiang *et al.* (2015) failed to express WW1 as a GST-tagged protein, suggesting that the use of a GB1-tag has increased the expression of this protein domain potentially enabling further analysis of this domain by NMR. However, as the backbone assignment and HSQC results suggest, unbound WW1 still remains troublesome for structural solution. One explanation though for having an incomplete folding pattern when in the absence of a substrate, may be that WW1 only folds completely upon binding of its substrates. Analysis of the closely related WW3 domain of Nedd4-1 found that in the absence of ligand, WW3 existed in both an unfolded state and a conformation primed for substrate binding (Panwalker *et al.*, 2016). Upon the addition of ligand however the only state detected was the completely folded ligand-bound WW3. Hence, addition of substrate may be able induce complete folding of WW1 enabling further research into the role and targeting of this domain.

Smad proteins 2, 3 and 7 are known to interact with several members of the Nedd4 family. While all WW domains of Nedd4L, Smurf 1 and Smurf 2 showed some kind of binding to Smad 7 peptide in ITC experiments (Aragón *et al.*, 2012), NMR analysis of WWP2 WW1 showed no significant binding of Smad 7 peptide in contrast to all other

WWP2 WW domains which have been shown to interact with Smad 7 peptide. This suggests that WW1 is unique within WWP2 in its ability to solely bind to R-Smads instead of I-Smad 7 as well. As for WW1 interactions with Smad 2 and 3 peptides, changes in 2-3 residues were observed suggesting either the start of direct binding of Smad 2 and 3 peptides to the domain or that WW1 was beginning to fold. Due to a lack of saturation however the exact nature of the interaction could not be completely deduced. The changes in residues 322 Asn (87 Asn in GB1-WW1 and GB1-WW1-2) and 327 Thr (92 Thr) in the Smad 2 titration and 322 Asn, 323 Thr (88 Thr) and 327 Thr in the Smad 3 titration suggest a very similar initial change in WW1 in the presence of these peptides. This is most likely due to single amino acid difference, hence further exploration of this at higher saturation points is required to determine whether these peptides do interact with slightly different residues or in different orientations. The minor observations in these titrations that appear to indicate the start of binding contradict previous work which suggested that Smad 2 and 3 had to be phosphorylated at least at T179 by CDK for interaction with Nedd4 E3 ligases (Aragón *et al.*, 2011; Gao *et al.*, 2009; Wang *et al.*, 2009). As phosphorylation of these proteins plays a major role in regulating their secondary messenger activities, this is something that requires further exploration to determine whether WWP2 could be a phosphorylation-independent E3 ligase for these proteins. Additionally, determining whether the presence of a phosphate group influences the side chains that interact may unveil another level of specificity that could be targeted via WW domain interaction surface mimics. Interestingly, using Jpred 4 (Drozdetskiy *et al.*, 2015; Cole *et al.*, 2008) for structural prediction of WW1, to get an idea of the location of the residues 322, 327 and 323, suggested that they reside in the loop region after the second β -strand and the third β -strand. While other Nedd4 E3 ligases have shown the involvement of residues in loop regions, this may suggest that the changing chemical shift positions may be more likely due to allosteric conformational change induced by the presence of peptide rather than direct binding to these residues, but as mentioned earlier more work is needed.

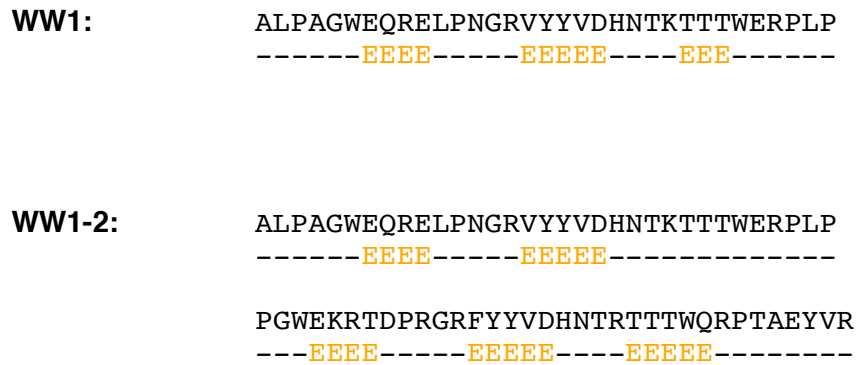


Figure 4.3.1 - Jpred 4 simplified representation of WW1 and WW1-2 predicted structures.

– refers to flexible regions; E represents a β -strand region.

In comparison to WW1, the structural predictions for WW1-2 construct presented 5 not 6 β -strands (Figure 4.3.1). This seems to suggest that the third β -strand of WW1 might fail to form and instead may remain flexible producing a “linker” region between the two domains. In doing so, this may be a mechanism of regulation by WWP2-FL in controlling interactions with substrates. As briefly mentioned previously, tandem domain cooperativity is gaining increasing evidence, suggesting this as a possible regulator of WW domain activity of Ned4 E3 ligases (Aragón *et al.*, 2012; Chong *et al.*, 2010; Dodson *et al.*, 2015; Fedoroff *et al.*, 2004). To further understand this region of WWP2-FL, the tandem domain GB1-WW1-2 was also used in titration analysis. Results showed significant differences to the WW1 only results, suggesting that the presence of WW2 influences the folding and binding of WW1 to substrates. Due to only partial assignments at this stage, it is difficult to determine which residues across both domains are able to interact with the Smad peptides. However, changes in 335 Gly (100 Gly in GB1-WW1-2), 343 Arg (108Arg) and 354 Arg (119 Arg) suggest WW2 can bind Smad 2 and 3 peptides, which has yet to be analysed. As to whether a single moiety of Smad 2/3 binds to both domains or whether one domain binds and folds the other has yet to be elucidated. As many peaks are seen to undergo perturbation during the titration, this supports these domains being able to interact without the presence of phosphorylation, although this modification may change the interactions. The tandem domain seemed to also undergo folding upon interaction with the peptides, reflecting a previously mentioned argument that this region of WWP2 may only fold upon substrate binding in an induced fit mechanism of action. Analysis of the Smad 7 interaction with WW1-2 also highlighted folding upon Smad 7 interaction, although this is hypothesised to be from binding of the second WW domain to Smad 7 peptide inducing folding of WW2 which in turn induced WW1 folding. It would be interesting to determine whether

this is the case or if the binding of Smad 7 to WW2 forces an interaction between WW1 and the peptide. Due to the inability to transfer all assignments from a fully assigned Smad7 bound GB1-WW2, this suggests WW2 within WW1-2 is in a slightly different conformation when WW1 is present therefore WW1 does have some influence on WW2 when its adjacent. Given the 5 different models for cooperative activity of tandem domains summarised by Dodson *et al.* (2015), in conjunction with the lack of a clear linker region between WW1 and WW2 in the full-length protein, tandem domain WW1-2 requires further investigation as to what role each domain plays when present together, and to structurally determine the bound states of these to potentially uncover the structural basis of the mechanism of action as well as the key residues on the interaction interface.

5. High-throughput screening of compounds for WWP2 inhibitor discovery

5.1. Introduction

WWP2 resides in the Nedd4 family of HECT E3 ligases (Pirozzi *et al.*, 1997). This ligase is known to interact with tumour suppressors such as Smads 2, 3 and 7 in the TGF- β signalling pathway (Soond and Chantry, 2011a; Soond and Chantry, 2011b) and PTEN, a negative regulator of AKT signalling (Maddika *et al.*, 2011). Ubiquitination and in turn proteasomal degradation of these substrates by WWP2 is believed to be vital to the regulation of these signalling pathways. Dysfunction in the expression of WWP2, and its close relative WWP1, has been linked to the development and progression of cancer (Soond *et al.*, 2013; Chen *et al.*, 2006). Previous work by Soond and Chantry (2013) suggests WWP2 overexpression leads to decreased availability of inhibitory Smad 7 and stabilisation in the level of Smad 3, leading to cancer cell survival and EMT. Using the knowledge of how these ligases target and modify tumour suppressors, novel therapies can be developed for use in cancer treatment. In the previous chapter, the use of NMR titrations for interaction analysis was discussed with the future potential for inhibitory peptide design. While peptide-based inhibitors have shown high selectivity and efficacy, there are many difficulties yet to be overcome such as drug delivery, peptide half-life, and the ability to migrate across the cell membrane (Lau and Dunn, 2018). One way to overcome some of these disadvantages is to use small molecules. This chapter focuses on small molecule screening using a biochemical assay as an alternative approach to the discovery of WWP2 inhibitors.

Small molecule inhibitors of Nedd4 E3 ligases are becoming of interest as a focus for research into novel therapies. However, due to the complex nature of the ubiquitination cascade very few have been published to date. These enzymes rely on protein-protein interactions and thioester-based reactions to function, making them very difficult to selectively target. The first E3 ligase inhibitors to be discovered were the nutlins, which inhibit the RING E3 ligase Mdm2 preventing interaction with p53, a well-known tumour suppressor protein (Vassilev *et al.*, 2004). This was the first evidence that inhibition of these ligases was possible using small molecules. This then led to further research into whether this could be implemented for HECT E3 ligases as well. In 2014, a small molecule inhibitor known as heclin was discovered (Mund *et al.*, 2014). Heclin showed inhibition of a variety of Nedd4 E3 ligases with IC₅₀ values in the micromolar range, suggesting good effectiveness at inhibiting members of the Nedd4 family. Similarly, the FDA approved drug clomipramine was shown to inhibit Itch (Rossi *et al.*, 2014). However, both studies highlighted a lack of selectivity in targeting a specific Nedd4 E3 ligase over others. This evidence suggests that while it is possible to use small

molecules to inhibit Nedd4 E3 ligases, they currently lack the specificity towards individual HECT E3 ligases. Further investigation into inhibitor discovery against Nedd4 E3 ligases has identified additional compounds capable of reducing ubiquitination activity. An active anti-proliferative compound found in cruciferous vegetables called indole-3-carbinol (IC3) stabilised PTEN levels in melanoma cells by preventing the ubiquitination and degradation of PTEN mediated by Nedd4 (Aronchik *et al.*, 2014). Development upon this work with analogues of IC3 showed these compounds to be very effective against Nedd4 and directly binds to Nedd4, supporting these compounds as a new group of Nedd4 inhibitors (Quirit *et al.*, 2017). *In silico* modelling predicted the binding site to reside within the HECT domain, similar to the other inhibitors currently discovered, and be selective towards Nedd4 based on HECT domain sequence similarities and absence of a similar predicted binding site within the other ligases. As other ligases were not experimentally tested, it would be interesting to see conclusively whether these are specific to Nedd4 or can inhibit other family members. Other research has also identified a covalent inhibitor of Nedd4 HECT activity (Kathman *et al.*, 2015). Compound 1 in this work was found to interact with a non-catalytic cysteine residue in the HECT domain. The resultant effect was a switch in activity from processive polyubiquitin chain formation, where only one E3~substrate binding event is required for polyubiquitination, to a distributive in which multiple encounters are required for polyubiquitin chain formation on the substrate. This highlighted that not only is the catalytic cysteine residue a target for inhibition but other areas of the HECT domain may also be manipulated in drug design to alter the activities of Nedd4 E3 ligases. Furthermore, the selectivity of compound 1 appeared to be towards ligases which contained this specific cysteine residue as no interaction was observed with E6-AP and WWP1, suggesting that inhibitor design towards specific Nedd4 E3 ligases is possible with further understanding of the mechanisms and differences in both structure and function of these ligases.

The discovery of the small molecule inhibitors commonly uses compound libraries, as described by Rossi *et al.* (2014). These libraries can consist of compounds with a wide range of structures and functional groups, fragment-based sets or even approved FDA therapies. Choosing the right set of compounds is important for the protein of interest and assay type being used. This method provides a quick and efficient screen for initial inhibitory hit compounds. Any hits identified then undergo further testing and modification to increase the specificity and inhibitory effect. There are many types of assays used for compound screening, such as virtual screening, biochemical assay

screening and compound binding assays. Unlike virtual screening and binding assays, the use of a biochemical assay for compound screening can give more biologically relevant results as to whether a compound is capable of inhibiting the target protein activity.

The main aim of this chapter is to use small molecule screening as an alternative approach to NMR-based peptide design. The compound screening in this chapter utilised small molecules in an enzymatic plate assay using auto-ubiquitination as a readout to find initial hits. Compound libraries used in this work were chosen on the basis of utilising the continual refinement of the NIH NCI compound sets that aims to reduce potential interference/non-specific compound activity, while retaining a wide variety of molecule structures. In addition to the diversity set and natural products, the oncology set was also screened to determine whether any current FDA approved cancer therapies are active inhibitors of the ubiquitination mechanism. Positive hits for inhibition were then further tested in triplicate and in dose-dependent assays to determine the effectiveness of these compounds. The second aim is to determine whether the compounds selectively inhibit WWP2. This was carried out using ubiquitination assays against UbcH7, Uba1, Nedd4 and WWP1. The final aim is to determine whether these compounds had a significant effect on the migration of prostate cancer cells. Malignant prostate cancer cell lines PC-3 and DU-145 were chosen for these experiments due to both being androgen insensitive and TGF- β sensitive. This was exploited to reduce the likelihood of any effects being due to androgen-TGF- β signalling pathway cross-talk. Additionally, while DU145 are PTEN positive, PC-3 are PTEN negative thus providing alternative pathway statuses for comparison when testing the effect of compounds.

5.2. Results

5.2.1. Auto-ubiquitination biochemical assay optimisation

A truncated isoform of WWP2 (WWP2-C) which exists within cells was initially used for compound screen assay set-up. To determine whether WWP2-C can be used in an auto-ubiquitination assay, GST-tagged WWP2-C and WWP2-C^{C399A}, a mutant version where the catalytic cysteine in the HECT domain was replaced with alanine, were purified on glutathione coated beads. Reactions were set up as described in section 2.2.9.1, with the appropriate controls as shown in Figure 5.2.1. Samples were analysed by western blot using an anti-HRP-streptavidin conjugated antibody to detect biotinylated-ubiquitin. Figure 5.2.1A shows the presence of poly-ubiquitin chains and ubiquitin-bound higher molecular weight proteins with GST-WWP2-C but not GST-WWP2-C^{C399A}. This indicates that WWP2-C is capable of auto-ubiquitination, and that this activity is vastly reduced by mutation of the catalytic cysteine residue, which could be used as a low activity control. The presence of ubiquitin charged E1 with the mutant supports the difference in ubiquitination levels being due to WWP2-C and not earlier in the ubiquitination mechanism.

The next step was to determine whether GST-WWP2-C auto-ubiquitination could be inhibited using a known inhibitor of Nedd4. Reactions were set up as described in section 2.2.9.1 with heclin added in a dose-dependent manner. Samples were analysed by SDS-PAGE and western blotting against biotinylated ubiquitin. A noticeable decrease in auto-ubiquitination was observed at 50 μ M and 100 μ M (Figure 5.2.1B) compared to the control, suggesting that heclin inhibits WWP2-C. This supports the use of WWP2-C for inhibitor discovery, however prevents the use of heclin as a positive control for inhibition due to the high concentration required for inhibition to be detected.

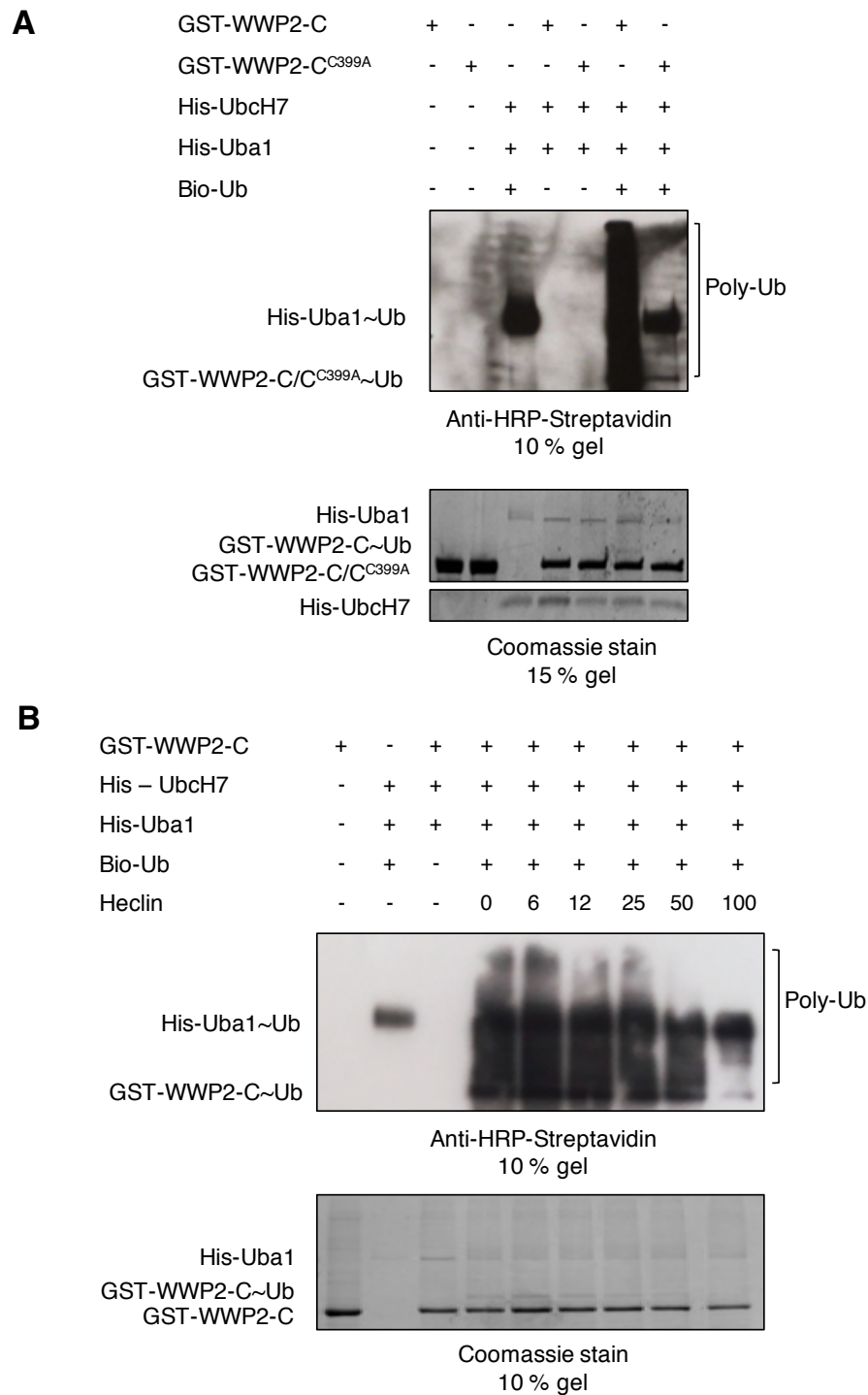


Figure 5.2.1 - GST-WWP2-C auto-ubiquitination and inhibition test.

GST-WWP2-C/-C^{C399A} coated glutathione beads were incubated with components described in the figure and samples analysed by SDS-PAGE and western blot using anti-HRP-streptavidin against biotinylated ubiquitin (Bio-Ub). UbcH7 – E2 conjugating enzyme; Uba1 – E1 activating enzyme. A) Auto-ubiquitination assay using GST-WWP2-C and GST-WWP2-C^{C399A} beads incubated with components described and analysed by SDS-PAGE and western blot against HRP-streptavidin (n = 1). B) Dose-dependent GST-WWP2-C auto-ubiquitination assay using Heclin (n = 1).

Similar reactions were set up in 96 well format on glutathione-coated plates (adapted from Rossi *et al.*, 2014) to see whether WWP2-C auto-ubiquitination could also occur in plate format as a high through-put screening method. The plate was initially incubated with cell lysate to immobilise GST-WWP2-C. Reactions were set-up as described in section 2.2.9.2, and auto-ubiquitination detected using anti-HRP-streptavidin and TMB. The TMB reaction was stopped using HCl. Absorbance was measured at 450 nm using a Hidex Sense microplate reader. There was no significant difference between WWP2-C and WWP2-C^{C339A} across all samples (Figure 5.2.2A). This indicates auto-ubiquitination did not occur, as all OD values fell below 0.25 with no significant difference compared to buffer alone. The assay was also tested using varying concentrations of components, fresh expression pellets, an additional BSA blocking step and FLAG-tagged ubiquitin and anti-FLAG-HRP conjugate to determine if the problem was the assay quantities and set-up, however similar results were observed (data not shown). Western analysis showed binding of GST-WWP2-C to the microplates, however no laddering was detected, indicating the problem was the reaction (Figure 5.2.2B). This suggests that the level of WWP2-C auto-ubiquitination is too low, if present at all when immobilised, for use in this assay. This is potentially due to the limited number of ubiquitination sites within WWP2-C which only consists of WW4 and the HECT domain.

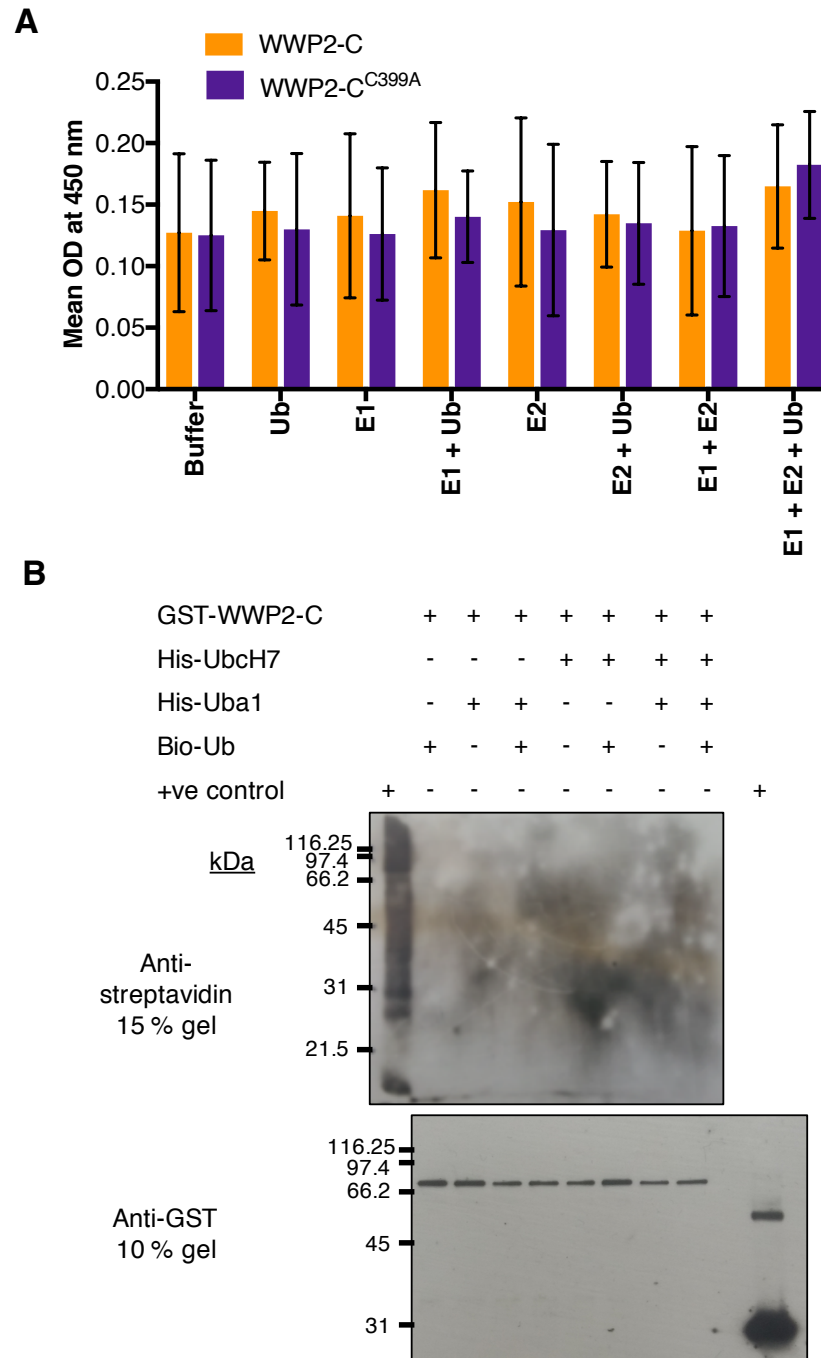


Figure 5.2.2 - GST-WWP2-C plate assay optimisation.

GST-WWP2-C and -C^{C399A} were incubated with E1 (His-Uba1), E2 (His-UbcH7) and biotinylated ubiquitin (Bio-Ub). A) After plate washing, anti-HRP(horse-radish peroxidase)-streptavidin and TMB (3,3',5,5'-tetramethylbenzidine) were used to detect and quantify auto-ubiquitination levels. Data shown is from 3 independent experiments (n=3). Error bars show SD; statistical analysis using two-way ANOVA and Tukey's multiple comparison across all means, $P > 0.1$. B) Western blot analysis of microplate samples after plate washing, using anti-HRP-streptavidin and anti-GST antibodies (n = 1). Data representative of 3 repeats.

Due to WWP2-C producing no significant readout, full-length WWP2 protein was chosen for the microplate assay. GST-tagged WWP2-FL and GST-WWP2-FL^{C838A} were synthesised using phusion and in-fusion cloning outlined in section 2.1.2 and 2.1.4, followed by bacterial expression as described in section 2.2.1. An optimisation test was carried out using the same quantities and protocol as outlined in section 2.2.9.2, using the appropriate controls as seen in Figure 5.2.3. Results showed a significant increase in activity when E1, E2 and Ub were present for WWP2-FL, however no significant effect was observed for WWP2-FL^{C838A}. Furthermore, all controls showed no significant increase in OD compared to buffer. This suggests that the plate assay set up was successful with WWP2-FL and can be used in further experiments. By mutating the catalytically active cysteine residue 838, there is no significant auto-ubiquitination detectable, showing that this residue is vital for detectable auto-ubiquitination levels.

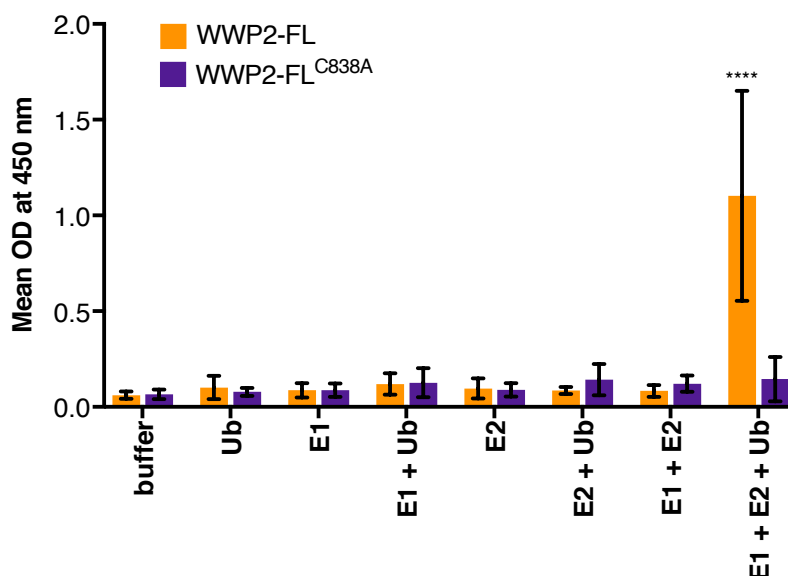


Figure 5.2.3 - GST-WWP2-FL plate assay optimisation.

GST-WWP2-FL and -FL^{C838A} were incubated with E1 (Uba1), E2 (UbcH7), Flag-ubiquitin (Ub) and ATP. Levels of auto-ubiquitination were detected and quantified using anti-Flag-HRP (horse-radish peroxidase) and TMB (3,3',5,5'-tetramethylbenzidine). The data shown is from 3 independent experiments (n=3). Error bars showing standard deviation; statistical analysis using two-way ANOVA with Tukey's multiple comparisons of all means to each other, ****P < 0.0001.

5.2.2. Initial screening of compound libraries

On the basis that GST-WWP2-FL provides a reliable output in 96 well format, compound screens were set up as described in section 2.2.9.2. A total of 2141 compounds were screened from NCI Oncology set II, Diversity set V and Natural Products set IV (NCI, US). Internal plate controls for 100% activity (DMSO – GST-WWP2-FL plus all reaction components and 0.1 % DMSO) and 0 % activity (E3 only – GST-WWP2-FL in ubiquitination buffer) were used to calculate the percentage residual activity of the compound wells. The initial screens were run as single shot assays, with Z' values over 0.5 suggesting that these plate assays were robust. The controls for each library screen (right panels in Figure 5.2.4) show no significant difference between the buffer and E3 only low controls. A significant increase was observed between the low and high controls, mirroring the results in Figure 5.2.3. No significant effect was observed in the presence of DMSO, suggesting any change in activity is not due to DMSO. Results were normalised using the high and low controls mentioned above to calculate the percentage residual activity compared to DMSO control. A threshold of 70 % (> 2 SD from the DMSO control mean) was used to determine which compounds would be selected for further testing. The data shows 24 hits from the diversity set (Figure 5.2.4A), no hits from the oncology set (Figure 5.2.4B) and 21 hits from the natural products (Figure 5.2.4C).

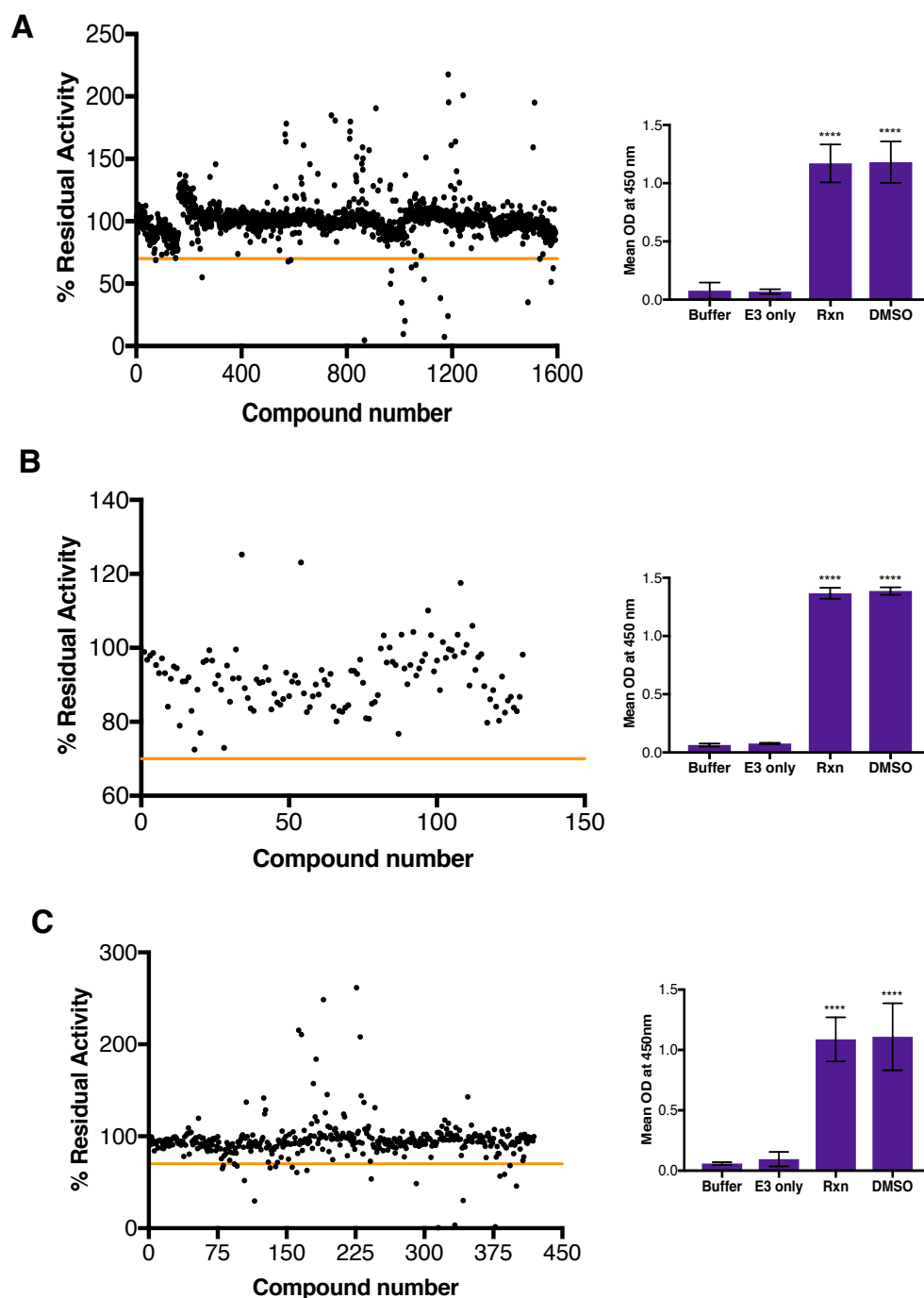


Figure 5.2.4 - Compound screens using GST-WWP2-FL auto-ubiquitination assay.

Assays were carried out using 10 μ M compound at 0.1 % DMSO in single shot with internal controls for 100 % activity with DMSO (DMSO) and without DMSO (Rxn), and 0 % activity (E3 only and Buffer) on each individual 96 well plate. Auto-ubiquitination levels were detected by anti-Flag-HRP (horse-radish peroxidase) conjugated antibody and TMB (3,3',5,5'-tetramethylbenzidine). OD was measured at 450 nm. A) Diversity set V screen and plate controls (right panel) (n=1). B) Oncology set II screen and plate controls (right panel) performed as described in A (n=1). C) Natural Products set and plate controls (right panel) performed as described in A (n=1). E3 only – GST-WWP2-FL in ubiquitination buffer; Rxn – GST-WWP2-FL plus all reaction components. DMSO – GST-WWP2-FL reaction with 0.1 %

DMSO. All assays were normalised using 100 % and 0 % controls and had a threshold of 70 % (—). $0.5 < Z' < 0.9$ for all single shot assays. Error bars showing SD; statistical analysis of controls using one-way ANOVA with Tukey's multiple comparisons of all means to each other, **** $P < 0.0001$.

5.2.3. Secondary screening of potential hits

Having identified several potential hits in the single shot assays, compounds below 70 % required further testing in triplicate to validate the inhibitory effect seen in the initial screening. This was carried out in the same manner as the single shot assay at 10 μ M and 0.1 % DMSO, however each compound was tested in triplicate. Figure 5.2.5A shows 18 compounds from the diversity set with a significant decrease in activity. Similarly, 10 compounds from the natural products set showed a significant decrease in activity. A threshold of 20 % and 50 % was set for the diversity set compounds (Figure 5.2.5A) and natural products (Figure 5.2.5B) respectively, to identify the best set of inhibitors to carry forward. 10 out of 18 diversity set compounds showed activity levels below the 20 % threshold, and 5 out of 10 hits from the natural products exhibited less than 50 %. This high stringency level allows only compounds showing greater inhibitory effect to be carried forward.

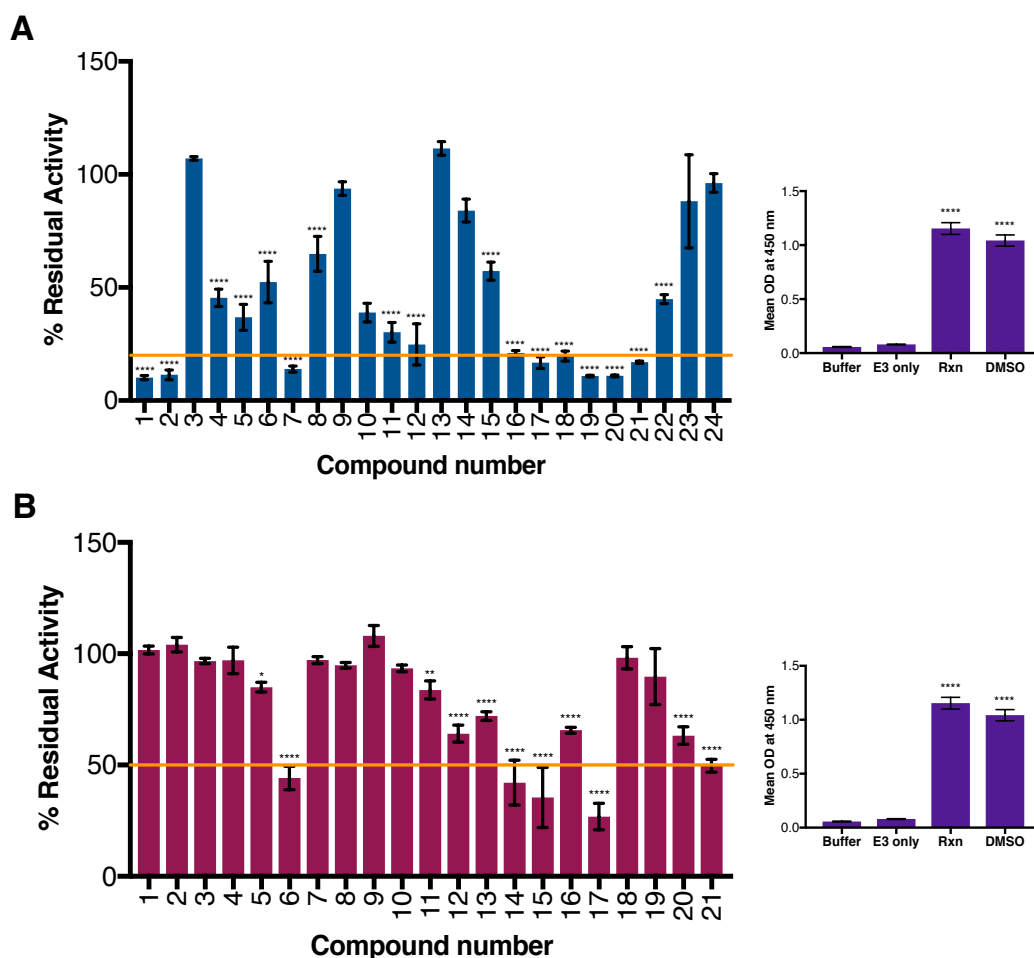


Figure 5.2.5 - Triplicate assays showing best potential hits from initial screen.

A) Auto-ubiquitination assay on 24 Diversity set V hits from section 5.2.2 at 10 μ M and 0.1 % DMSO with a threshold set at 20 % (—) (n=1). B) Natural product hits from section 5.2.2 at 10 μ M 0.1 % DMSO assayed in triplicate, with a threshold of 50 % (—) (n=1). E3 only – GST-WWP2-FL in ubiquitination buffer; Rxn – GST-WWP2-FL plus all reaction components. DMSO – GST-WWP2-FL reaction with 0.1 % DMSO. All data normalised using 100 % and 0 % activity controls. Plate controls are shown to the right. Error bars showing standard deviation; statistical analysis using one-way ANOVA with Dunnett's multiple comparison to DMSO control, * $P < 0.5$ and **** $P < 0.0001$, and Tukey's multiple comparison between all means (right panel), **** $P < 0.0001$.

5.2.4. Dose-dependent analysis of hit compounds

For these compounds to progress further as suitable hit compounds, a dose-dependent assay was used to assess the effectiveness of these compounds at inhibiting WWP2 auto-ubiquitination. This was also to derive \sim IC₅₀ values for more stringent refinement of the compounds. Concentrations from 1 – 100 μ M were tested at 1 % DMSO. Curves

were generated and IC₅₀ values calculated using non-linear regression of inhibitor concentration vs response. Results are shown in Figure 5.2.6, and IC₅₀ values with compound structures shown in Table 5.2.1. All compounds presented a dose-dependent inhibitory effect on WWP2 auto-ubiquitination, with IC₅₀ values ranging from 0.4 – 10 μ M. Compound 1 showed the highest effectiveness unlike compound 16 presenting the lowest effect across both sets of compounds. As all IC₅₀ values are ~10 μ M or below, this suggests these are all good starting compounds for further investigation and modification towards potential novel therapies.

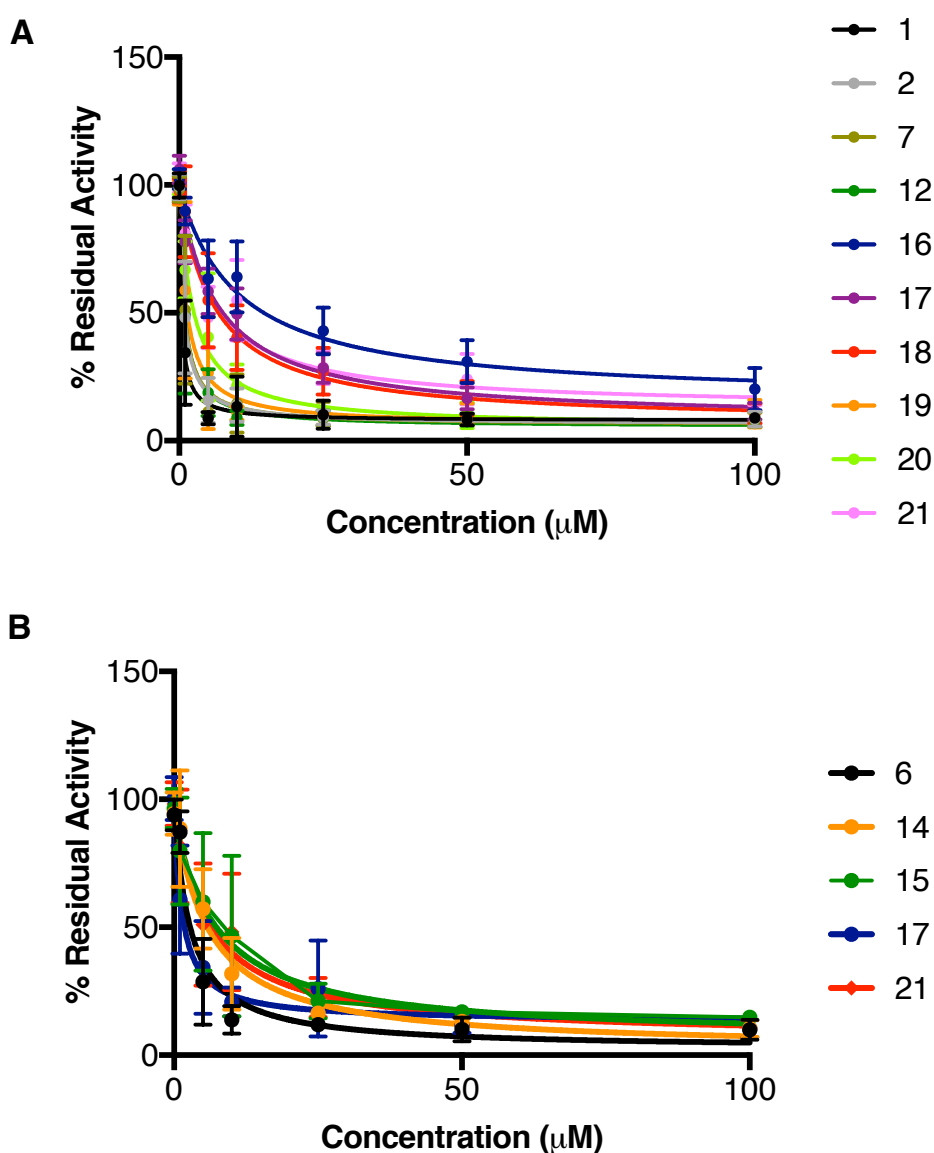


Figure 5.2.6 - Dose-dependent assays on 15 hits.

Auto-ubiquitination reactions used concentrations of compounds from 1 – 100 μ M at 1 % DMSO. A) Dose curves for Diversity set compounds. B) Dose curves obtained for natural products. Data shown is from three independent experiments ($n=3$). Error bars show standard deviation; statistical analysis using non-linear regression of concentration vs response, $R^2 > 0.8$.

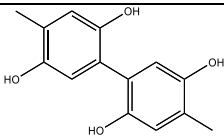
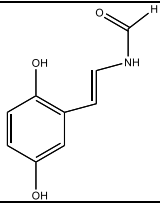
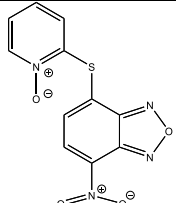
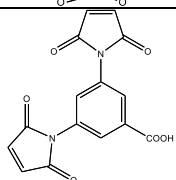
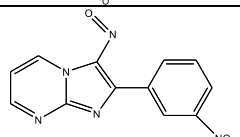
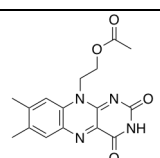
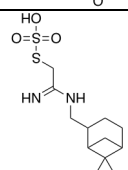
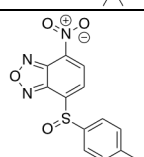
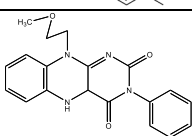
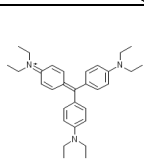
Number	NSC number	Structure	IC50 (μM)
1	2805		0.38
2	650438		0.77
7	228155		0.83
12	44750		0.85
16	369066		10.28
17	3064		6.51
18	194308		5.95
19	228150		1.29
20	288387		2.29
21	8675		5.39

Table 5.2.1 - Diversity set compound NSC codes, structures and IC50 values derived from dose-dependent analysis.

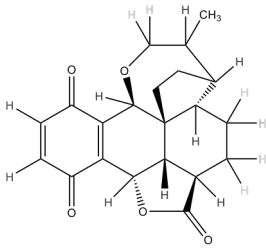
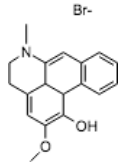
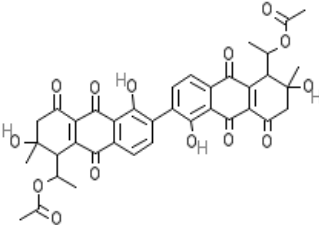
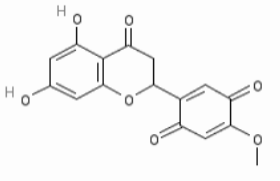
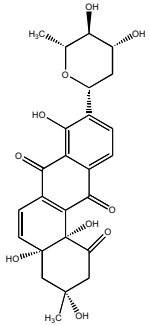
Number	NSC number	Structure	IC50 (μM)
6	401005		2.664
14	785168		5.736
15	248605		7.228
17	169517		1.337
21	184398		5.954

Table 5.2.2 - Natural products NSC codes, structures and IC50 values derived from dose-dependent analysis.

5.2.5. E2 and E1 counter screens

Previous results have highlighted 15 hits which can potentially proceed towards becoming lead compounds, however auto-ubiquitination is a complex reaction involving 2 other enzymes. Therefore, the next point of investigation was to test the specificity of

these hit compounds towards E3 ligases or whether they inhibit earlier in the auto-ubiquitination cascade. In order to determine the point of inhibition within the reaction pathway, ubiquitination assays were carried out using His-Ubch7 (His-E2) and GST-Uba1 (GST-E1) as described in sections 2.2.9.3 and 2.2.9.4, respectively. Initially, optimisation assays were carried to ensure that the assay could be used for compound analysis. These initial assays showed successful detection of ubiquitination of His-E2 and GST-E1 as shown in Figure 5.2.7. A significant increase in activity was seen when all reaction components were present compared to the controls. Furthermore, there was no significant difference between the absence and presence of DMSO at 0.1 % suggesting that the DMSO will not have an effect when testing the compounds. The overall OD output of the GST-E1 optimisation assay was much lower than for the His-E2 assay. This suggests that the productivity of GST-E1 may be reduced when immobilised or expressed as a fusion protein, or that the assay may require further optimisation, however as there was a significant increase in activity compared to the controls, the assay was used for compound testing.

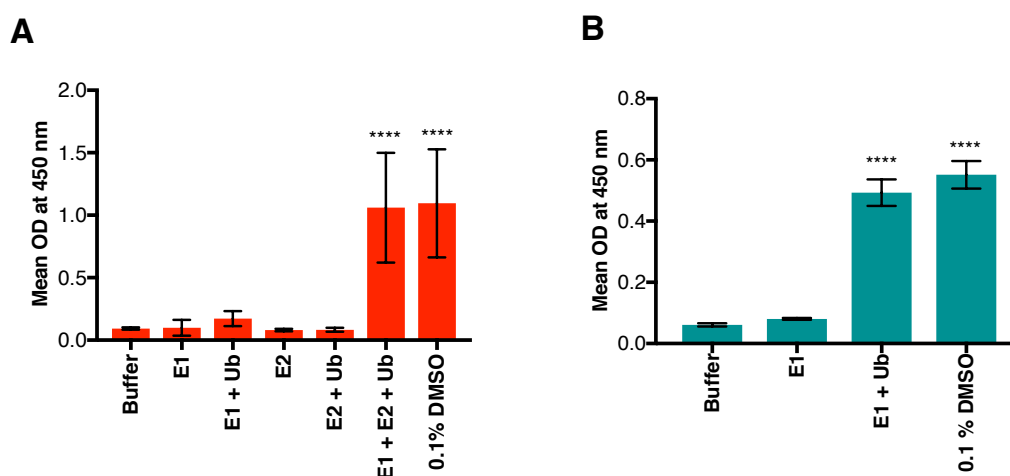


Figure 5.2.7 - His-Ubch7 and GST-Uba1 counter assay optimisations.

A) His-Ubch7 (E2) optimisation. E2 was coated onto nickel-coated plates and reactions containing E1, flag-ubiquitin (Ub) and ATP combinations were added. B) GST-Uba1 ubiquitination assay optimisation, using flag-ubiquitin (Ub). Auto-ubiquitination was detected using anti-Flag-HRP (horse-radish peroxidase) and TMB (3,3',5,5'-tetramethylbenzidine). The data shown is from three independent experiments (n=3). Error bars show standard deviation; statistical analysis using one-way ANOVA with Tukey's multiple comparisons of all mean values to each other, **** P < 0.0001.

Once the assay was optimised to obtain significant 100 % and 0 % activity controls, compounds were tested at the same initial concentrations as used in the GST-WWP2-FL screening, as described in sections 2.2.9.3 and 2.2.9.4. Only 4 of the 5 natural product hits could be tested due to availability of compounds. No significant decrease in activity was observed for all compounds in the His-E2 ubiquitination assay (Figure 5.2.8A), however all compounds in the GST-E1 assay showed significant decreases in activity levels compared to the DMSO control. All mean residual activity values were below the 70% threshold, suggesting all compounds have an inhibitory effect on GST-E1 in this assay.

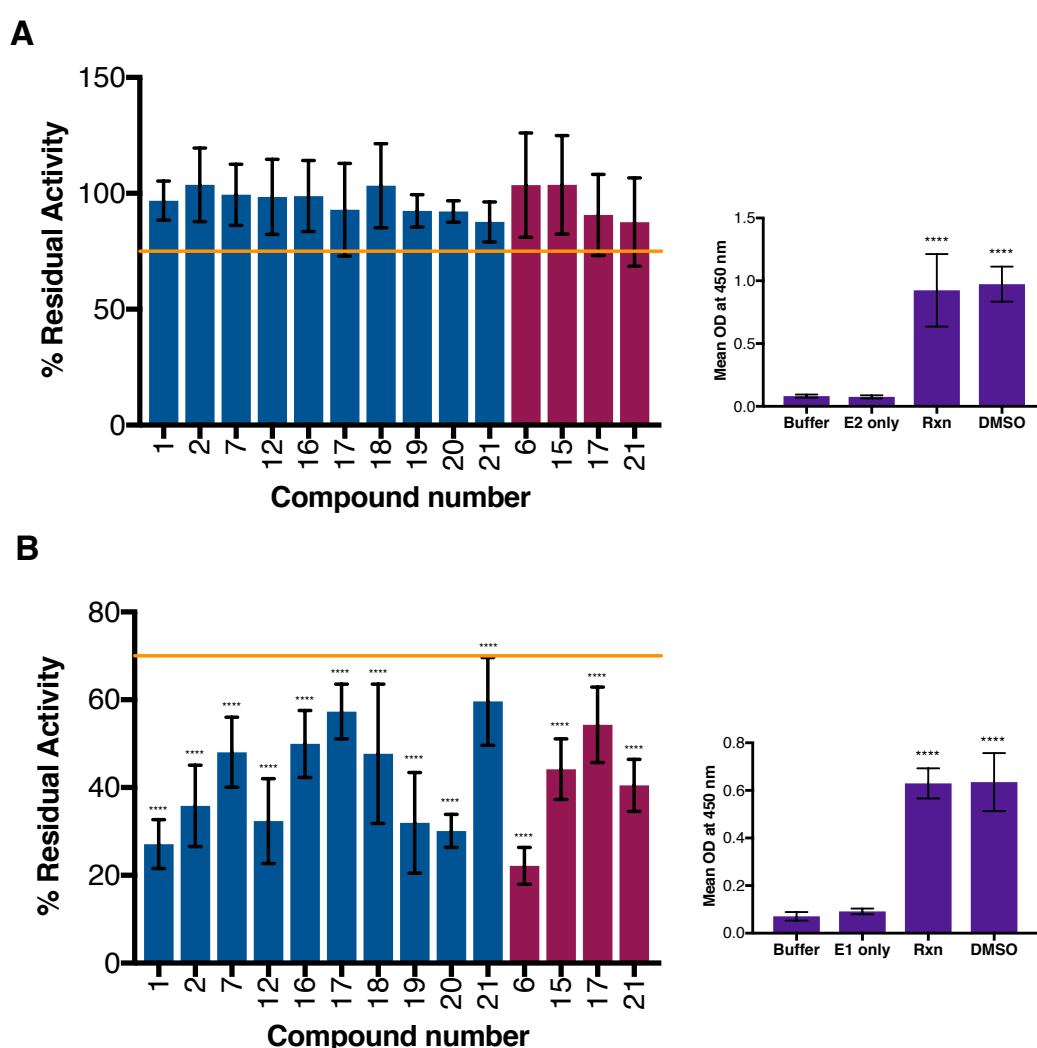


Figure 5.2.8 - Compound analysis against His-Ubch7 and GST-Uba1.

A) His-Ubch7 was immobilised on a nickel-coated plate and the assay run using GST-E1 and flag-ubiquitin. E2 only – His-Ubch7 and buffer; Rxn – His-Ubch7, GST-E1 and flag-ubiquitin; DMSO – His-Ubch7, GST-E1, flag-ubiquitin and 1 % DMSO. B) GST-Uba1 was coated onto a glutathione-coated plate, and flag-ubiquitin and ATP were added. E1 only – GST-E1 and buffer;

Rxn – GST-E1 and flag-ubiquitin; DMSO – GST-E1, flag-ubiquitin and 1 % DMSO. Ubiquitination was analysed by anti-flag-HRP (horse-radish peroxidase) and TMB (3,3',5,5'-tertamethylbenzidine) substrate. Blue bars showing diversity set compounds and red bars are natural products. Plate controls are shown to the right. The data shown is from three independent experiments (n=3). Thresholds were set at 70 % residual activity (—). Error bars showing standard deviation; statistical analysis using one-way ANOVA with Dunnett's multiple comparison with DMSO control, **** P < 0.0001, and Tukey's multiple comparison of all means to each other (right panel), **** P < 0.0001.

Due to the findings in the GST-E1 assay where all compounds showed a significant inhibitory effect on E1 activity, SDS-PAGE analysis of the GST-E1 reactions was used to validate these results. This was set up as described in section 2.2.9.5 and samples analysed by SDS-PAGE and coomassie staining. A single band was observed for His-E1 in the first lane of the DMSO control (Figure 5.2.9A and B). In the presence of ubiquitin, a second band appears at a higher molecular weight, as seen in lane 2, showing conjugation of ubiquitin to the E1 enzyme. This conjugation is prevented by the presence of β -mercaptoethanol as highlighted by the single band for His-E1 seen in lane 3. Comparatively, diversity set compounds 12, 16, 17, 20 and 21 all mirror this banding pattern showing no visible effect on E1 charging with ubiquitin (Figure 5.2.9A). It is unclear whether compound 1 had an effect on E1 activity because the bands were very faint, however the smearing in the second lane may suggest partial charging. Compounds 18 and 19 showed no clear separation of bands at the higher and lower molecular weight, making it unclear whether there has been E1-ubiquitin conjugation, however there may be some due the slightly higher position of banding in lane 2. The effect of compounds 2 and 7 were not determined due to time limitations, therefore require investigation to confirm activity on the E1 enzyme. As for the natural products shown in Figure 5.2.9B, all compounds show no significant effect on E1-ubiquitin complex formation as shown by the smearing in lane 2 which is not present in lanes 1 or 3. This suggests that while the compounds may be having a partial inhibitory effect on E1 activity as shown in the plate assay (Figure 5.2.8), there is no complete inhibition of the E1 enzyme as seen below in Figure 5.2.9 supporting the evidence in Figures 5.2.4 to 5.2.6 that the predominant inhibitory effect of these compounds is on E3 activity.

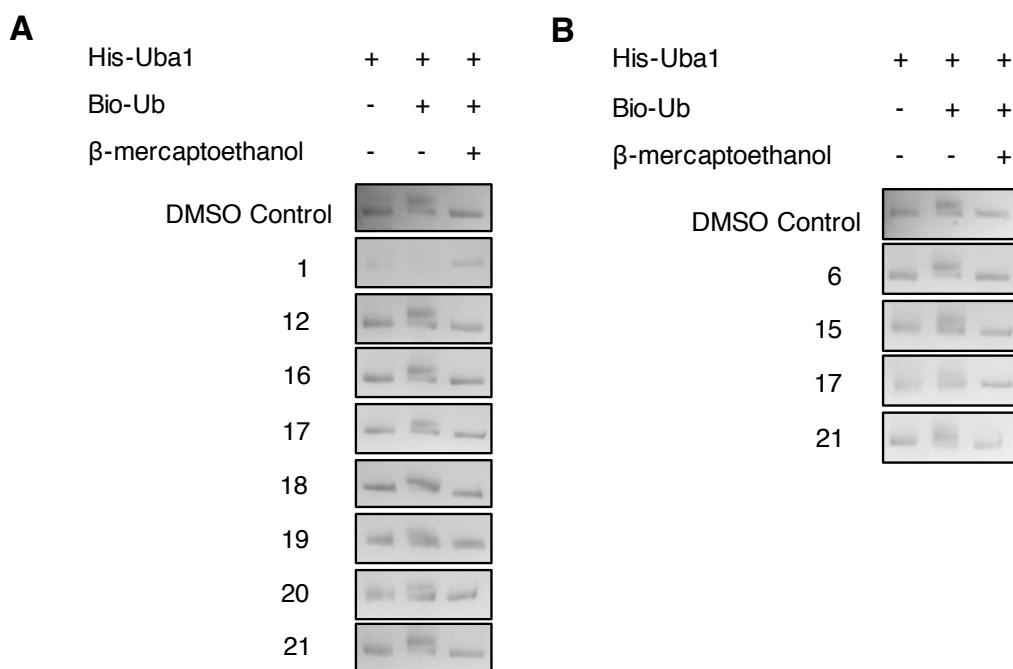


Figure 5.2.9 - His-Uba1-ubiquitin charging assay.

Purified His-Uba1 was incubated with either 0.1 % DMSO or 10 μM compound at 0.1 % DMSO in the presence or absence of ubiquitin (Bio-Ub) and 1 % β-mercaptoethanol, and samples analysed by SDS-PAGE and coomassie blue staining. A) E1 assay using Diversity set compounds (n = 1). B) E1 assay using Natural products (n = 1).

5.2.6. Nedd4 and WWP1 counter screens

Compounds were tested against Nedd4 and WWP1 to find out whether they were capable of inhibiting other HECT E3 ligases in the same family as WWP2. GST-tagged Nedd4 and His-tagged WWP1 were immobilised onto glutathione coated and nickel coated plates, respectively. Component reactions were prepared as described in section 2.2.9.6 and 2.2.9.7 and added to the appropriate wells. After incubation, auto-ubiquitination levels were determined using HRP-conjugated anti-Flag antibody and TMB substrate. Both Nedd4 and WWP1 produced a significant increase in OD when in the presence of E1, E2, and ubiquitin, unlike all other reactions (Figure 5.2.10). There was no significant difference between reactions which lacked one or more components, showing that Nedd4 and WWP1 are capable of auto-ubiquitination under these assay conditions, and provides both a 100 % and 0 % activity level which can be used to determine levels of inhibition by the 14 compounds.

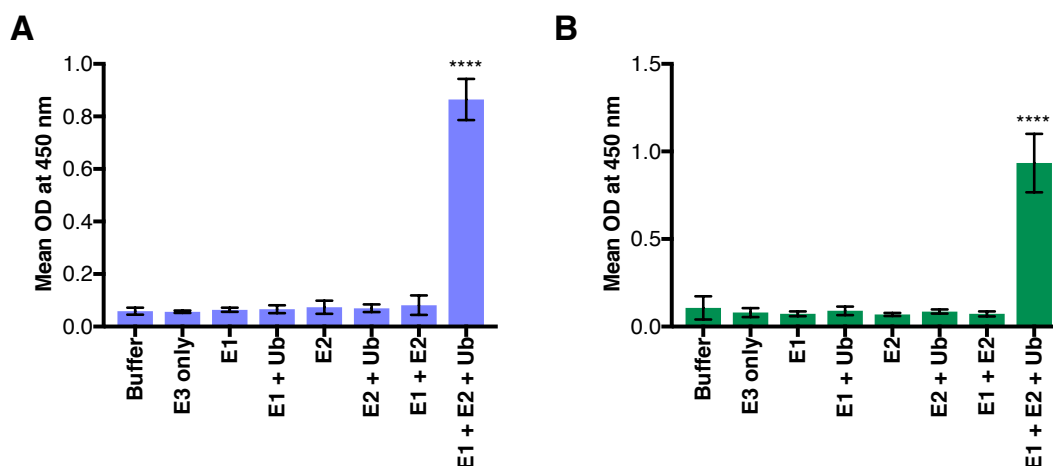


Figure 5.2.10 - HECT E3 ligase auto-ubiquitination counter screen optimisation.

A) GST-Nedd4 auto-ubiquitination optimisation. B) His-WWP1 auto-ubiquitination optimisation. HRP (horse-radish peroxidase) conjugated anti-Flag and TMB (3,3',5,5'-tetramethylbenzidine) were used to detect the presence of flag-ubiquitin after incubation of the coated plates with reaction components described above and in sections 2.2.9.6 and 2.2.9.7. E1 – Uba1; E2 – Ubch7; Ub – flag-ubiquitin. The data shown is from three independent experiments (n=3). Error bars show standard deviation; statistical analysis using one-way ANOVA with Tukey's multiple comparisons between all means, **** P < 0.0001.

To test the effect of the compounds on Nedd4 and WWP1 auto-ubiquitination, assays were set up as described in sections 2.2.9.6 and 2.2.9.7. When compared to the DMSO control, only natural product 17 showed no significant decrease in GST-Nedd4 activity (Figure 5.2.11A). All other compounds showed a significant reduction in auto-ubiquitination levels, with natural product 6 exhibiting the greatest level of inhibition. As for His-WWP1, all compounds produced a significant reduction in auto-ubiquitination levels (Figure 5.2.11B), in a similar pattern to the GST-Nedd4 assay. When comparing the 3 E3 ligase assays, some very clear differences between auto-ubiquitination levels were observed. Compounds 1, 2 and 12 from the diversity set and natural product 21 showed no significant difference between the HECT E3 ligases, implying that these compounds are not selective towards one of the ligases (Figure 5.2.11C). Nevertheless, all 4 do show a significant reduction in activity for all 3 proteins. Compounds 7, 19 and natural product 15 showed a significant difference between WWP2 and WWP1, and WWP1 and Nedd4 but 7 and 19 presented no difference between WWP2 and Nedd4. Natural product 15 showed a slightly higher % residual activity for Nedd4 than WWP2. This suggests a greater impact on WWP2 and Nedd4 activity over WWP1. Interestingly, compounds 16, 17, 18, 20 and 21 from the diversity set showed increased inhibition towards WWP2 over WWP1 and Nedd4, as

seen in Figure 5.2.11C. In contrast natural product 6 presented a greater effect on Nedd4 than WWP2 and WWP1. This implies these compounds show a relative degree of selectivity towards one of the 3 ligases analysed.

At this stage, the number of compounds for further work was reduced from 14 to 7 on the basis of how effective they were at inhibiting WWP2, therefore a threshold of 3 μ M IC50 value was used. Compound 2, cis-Erbstatin, was eliminated due to several reports of inhibitory activity by erbstatin against several other enzymes including tyrosine kinases (Imoto *et al.*, 1990; Markovits *et al.*, 1994), therefore the evidence acquired here may be due non-specific interactions rather than specific inhibition of the HECT E3 ligases. Also, compound 10, ethyl violet was removed from further work due to its established tumour cell toxicity properties via mitochondrial inhibition (Indig, 2002; Lewis and Goland, 1953) and use as a dye in histological techniques (Cong *et al.*, 2010).

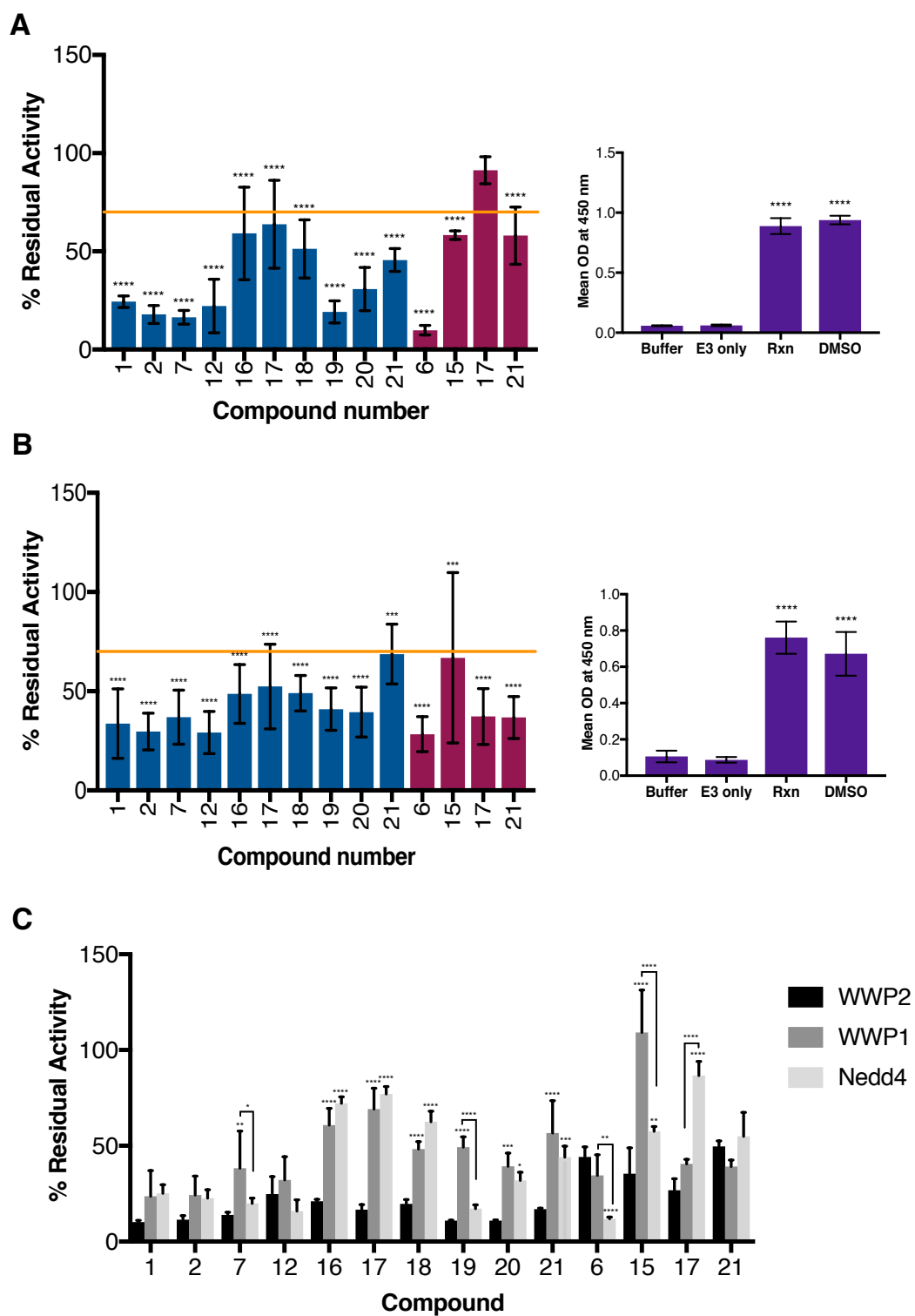


Figure 5.2.11 - GST-Nedd4 and His-WWP1 compound assays.

GST-Nedd4 and His-WWP1 were immobilised on glutathione- and nickel-coated plates respectively. Compounds were added followed by reaction components as described in section 2.2.9.6 and 2.2.9.7. Ubiquitin levels were quantified using anti-Flag-HRP (horse-radish peroxidase) and TMB (3,3',5,5'-tetramethylbenzidine), then read at 450 nm. A) GST-Nedd4 and B) His-WWP1 auto-ubiquitination assays using compounds at 10 μ M and 0.1 % DMSO. Blue bars showing diversity set compounds and red bars are natural products. Plate controls are

shown to the right. E3 only – Nedd4/WWP1 and buffer; Rxn – Nedd4/WWP1, UbcH7, Uba1 and flag-ubiquitin; DMSO - Nedd4/WWP1, UbcH7, Uba1, flag-ubiquitin and 1 % DMSO. The data shown is from three independent experiments (n=3). Thresholds were set at 70 % residual activity (—). C) Comparison of WWP2, WWP1 and Nedd4 compound inhibition (n=3). Error bars showing standard deviation; statistical analysis using one-way ANOVA with Dunnett's multiple comparison with DMSO control and Tukey's multiple comparison of all means to each other (right panel), **** P < 0.0001, *** P < 0.001, ** P < 0.01, * P < 0.05.

5.2.7. Compound inhibition of substrate ubiquitination

The previous data indicates that all 14 hit compounds, listed in Tables 5.2.1 and 5.2.2 are able to inhibit auto-ubiquitination of WWP2-FL, Nedd4 and WWP1 to varying degrees, with no effect on E2-conjugating and partial effect on E1-activating enzyme. However, this data doesn't indicate whether these compounds have the same effect on substrate ubiquitination. The next aim was to determine whether these 14 compounds could inhibit ubiquitination using GST-PTEN as a substrate. Reactions were composed of E1, E2, WWP2, PTEN, ubiquitin and ATP as described in section 2.2.9.8 and were analysed by western blotting using anti-GST, anti-Flag and anti-His antibodies. A reaction lacking WWP2 was used as a negative control (Figure 5.2.12A), and a reaction lacking E1 (Figure 5.2.12B). The presence of high molecular weight (MW) band in Figure 5.2.11A lane 3 suggests substrate ubiquitination, due to the absence of such banding when either PTEN or WWP2-FL are absent. This supports the band being a product of substrate ubiquitination and not WWP2-FL auto-ubiquitination. Figure 5.2.12B suggests all compounds except natural product compound 6 have an inhibitory effect on PTEN poly-ubiquitination, due to the absence of the high MW band present in the positive control. Compound 6 showed a faint high MW band suggesting reduced ubiquitination but not complete lack of. Collectively, this evidence supports the progression of these compounds as initial hit inhibitors for modification and further investigation into their binding and effectiveness against WWP2 activity.

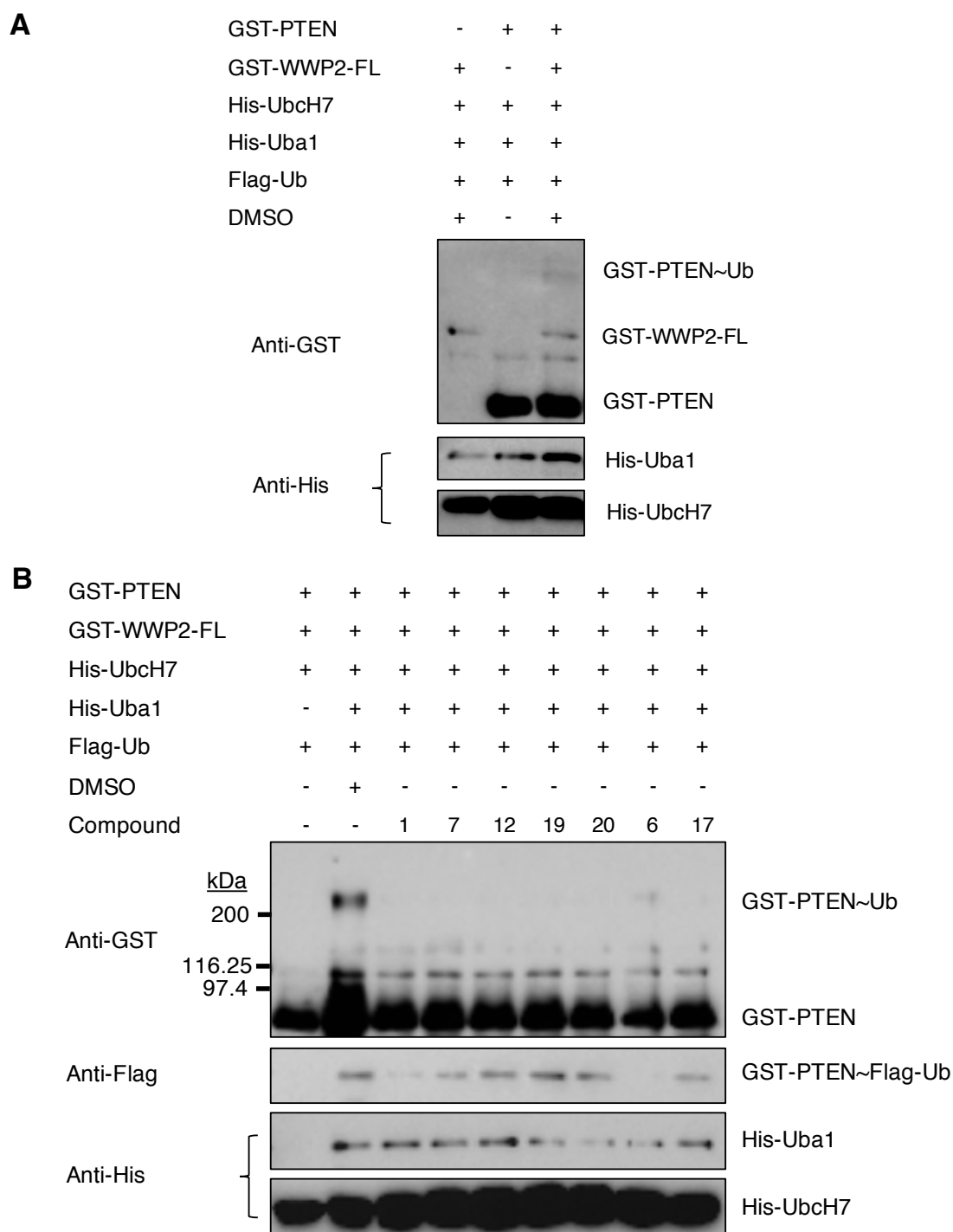


Figure 5.2.12 - GST-PTEN ubiquitination inhibition by compounds.

In vitro GST-PTEN ubiquitination reactions in the absence and presence of compounds, analysed using western blotting with anti-GST, anti-Flag and anti-His antibodies (n=1). The reactions are outlined in section 2.2.9.8, and incubated for 2 hrs. Reactions were stopped by boiling with laemmli buffer for 3 min. The negative control lacked WWP2 (A) or E1 (B). UbcH7 – E2 enzyme; Uba1 – E1 enzyme; Ub – ubiquitin.

5.2.8. Effect of compounds on advanced stage prostate cancer cell growth and migration

This section investigates whether any of the compounds highlighted so far inhibit the wound healing capabilities of advanced prostate cancer (PCa) cell lines. Due to one of the major steps in cancer cell survival and growth involving TGF- β signalling imbalance by WWP2 overexpression, discussed in chapter 1, scratch assays in the absence and presence of TGF- β were used to determine whether any of the small molecules could reduce cell growth/migration. PC-3 and DU145 cell lines were used as they are both metastatic PCa cell lines and present insensitivity to hormone treatment, a common characteristic of advanced prostate cancer (Cunningham and You, 2015; Kaighn *et al.*, 1979; Stone *et al.*, 1978). Cell lines were grown to 100% confluency and a scratch made in the monolayer. The assay was set up as described in section 2.3.4. Images were taken using phase microscopy at the time points shown below in Figures 5.2.13 - 5.2.20. Results were quantified to enable statistical analysis. Compound 7 showed a significant reduction in wound healing at 44 hr, compared to the DMSO control (Figure 5.2.13). This is supported by no significant difference between the control and DMSO (Figure 5.2.13B right panel), showing no effect by the presence of DMSO. All other compounds showed no significant change in cell growth/migration at all time points.

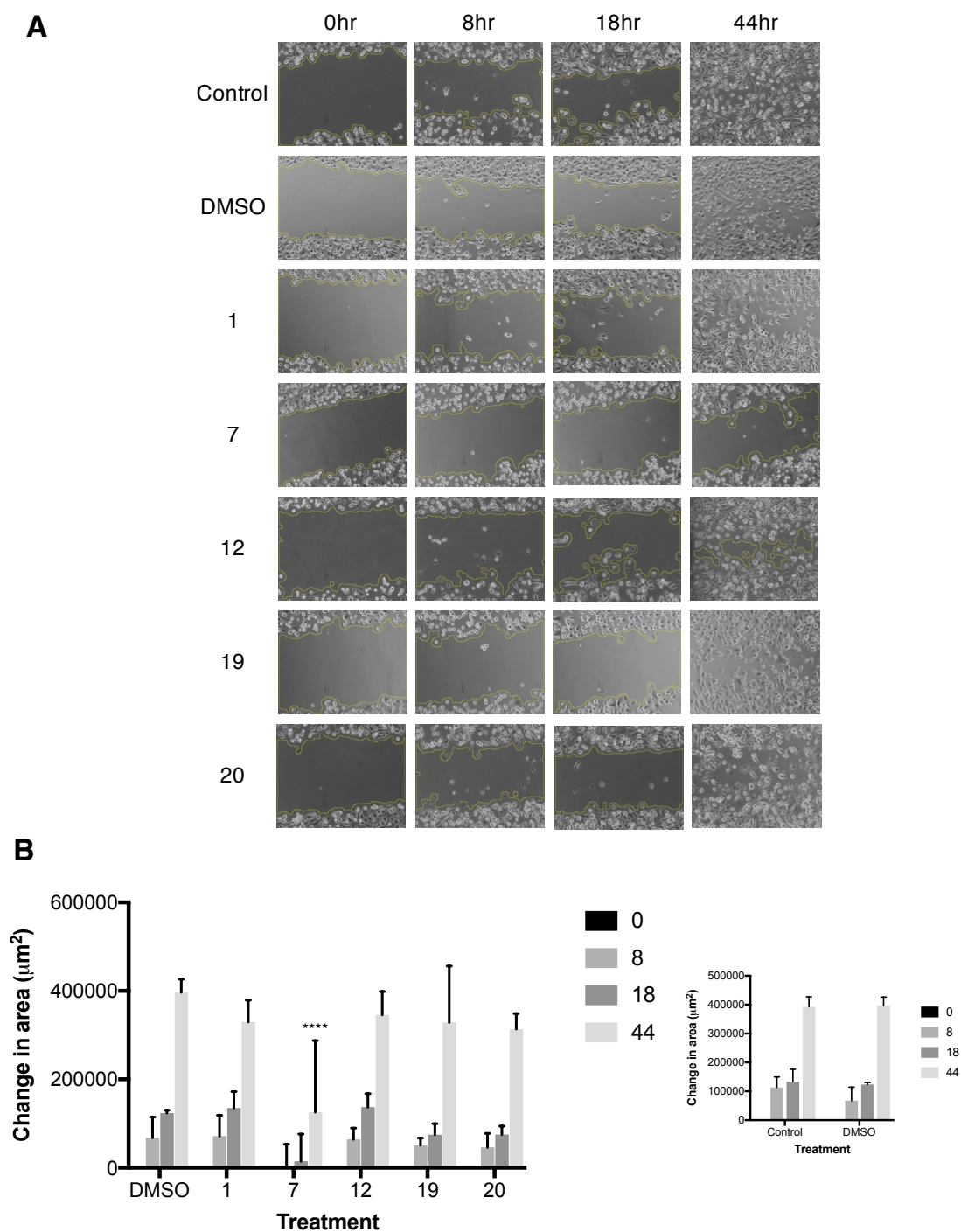


Figure 5.2.13 - PC-3 scratch assays with diversity set compounds in the absence of TGF- β stimulation.

PC-3 cells were grown to 100% confluency then incubated for 1 hour in 5 % FBS medium. Wells were washed after scratches were made and fresh medium containing the appropriate conditions, described in section 2.3.4, was added to wells. A) Images taken at 0, 8, 18 and 44 hr time points. Data representative of 6 technical repeats (n=6). B) Quantification of the change in area of scratches; statistical analysis using two-way ANOVA, with Dunnett's (left panel) and Sidak's (right panel) multiple comparisons; **** P < 0.0001.

With TGF- β stimulation, compounds 1 and 19 showed a decrease in migration at 44 hr (Figure 5.2.14), suggesting that these compounds inhibit scratch healing during activation of the TGF- β pathway. As for compound 7, a change in morphology and adherence was seen. Initially, cells were well adhered to the surface and presented mostly flat and some elongated forms. After the addition of compound 7 the cells began to change shape and became suspended in the media rather than fixed to the surface. This resulted in the inability to quantify wound healing. The cellular changes suggest that the compound may have either killed the cells during stimulation or altered EMT/adherence pathways, however this requires further investigation. All other compounds had no significant influence on growth/migration (Figure 5.2.14).

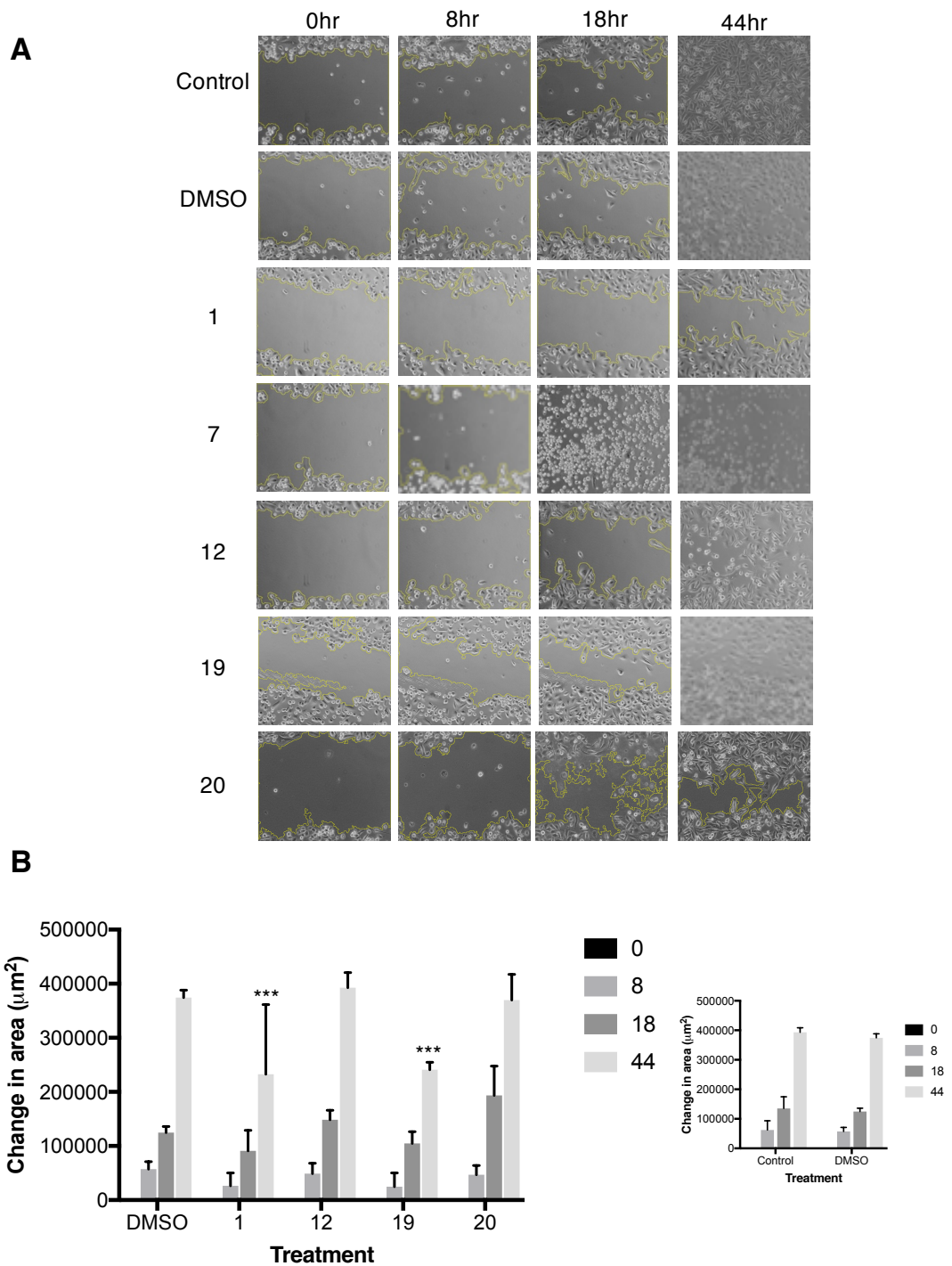


Figure 5.2.14 - PC-3 scratch assays under TGF- β positive conditions.

Once at 100 % confluency, PC-3 cells were incubated for 1 hour in 5 % FBS medium. Scratches were made, wells washed and fresh medium containing the appropriate conditions, described in section 2.3.4, was added to wells. A) Images taken at 0, 8, 18 and 44 hr time points. Data representative of 6 technical repeats (n=6). B) Quantification of scratch assay; statistical analysis using two-way ANOVA, with Dunnett's (left panel) and Sidak's (right panel) multiple comparisons; *** P < 0.001.

Results with the natural products 6 and 17 in the absence of TGF- β stimulation showed a significant decrease in wound healing at 44 hr with compound 6, but not 17 (Figure 5.2.15A and B left panel). No difference was observed between the control and DMSO (Figure 5.2.15B right panel), supporting inhibition of growth/migration by compound 6. In contrast, when stimulated with TGF- β a significant increase in wound healing was seen for DMSO (Figure 5.2.16B right panel), which was significantly reduced by the presence of compound 6 (Figure 5.2.16B left panel). Upon statistical analysis of the control compared to compound 6, no significant difference in wound healing was observed, suggesting compound 6 was able to inhibit the increased growth/migration seen with DMSO down to that observed for the control. This implies compound 6 may be showing some inhibition but would require further testing to confirm this due to the effect shown by DMSO in this assay.

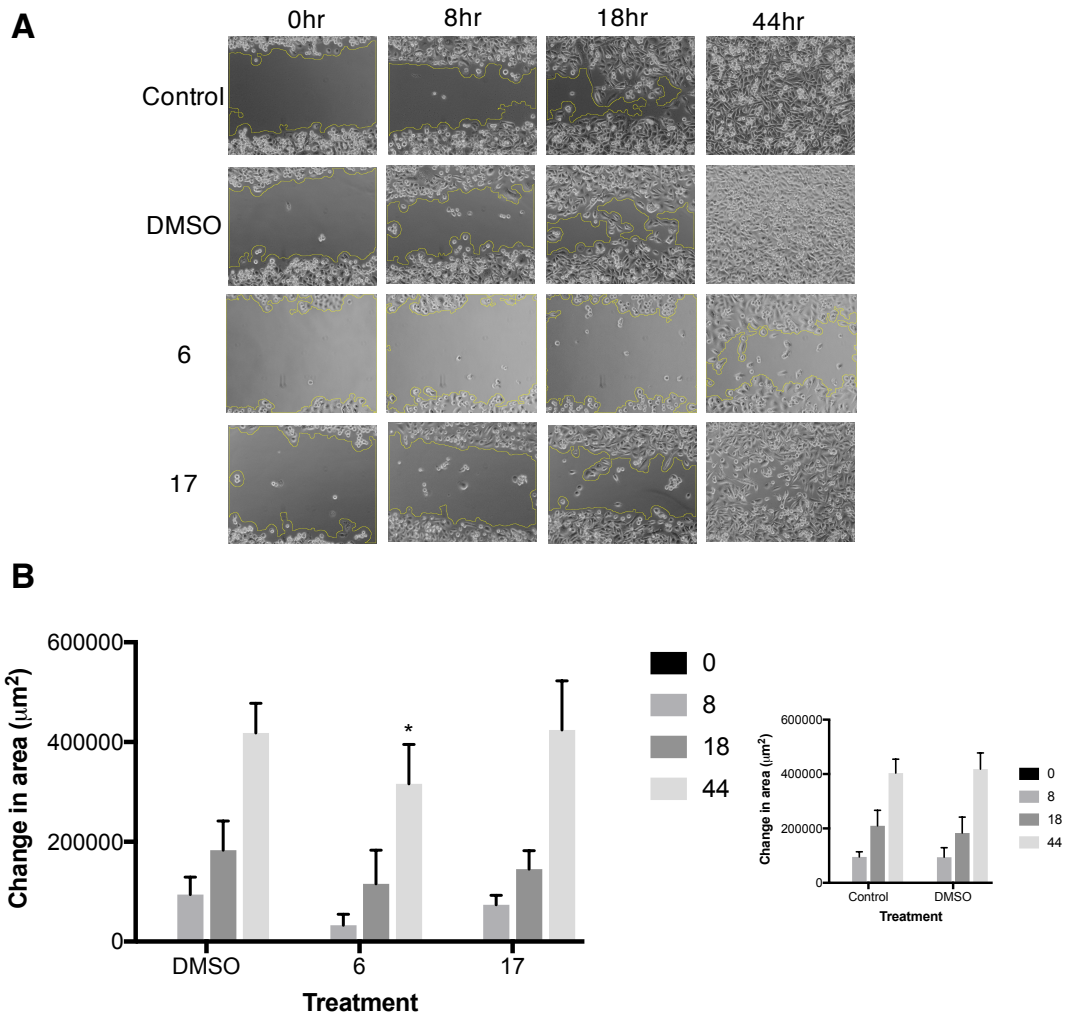


Figure 5.2.15 - Inhibition of PC-3 wound healing with natural products without TGF- β stimulation.

At 100% confluency, PC-3 cells were incubated for 1 hour in 5 % FBS medium. Scratches were made, wells washed and fresh medium containing the appropriate conditions, described in section 2.3.4, added to wells. A) Images taken at 0, 8, 18 and 44 hr time points. Data representative of 6 technical repeats (n=6). B) Quantification of the change in area of scratches; statistical analysis using two-way ANOVA, with Dunnett's (left panel) and Sidak's (right panel) multiple comparisons; * P < 0.05.

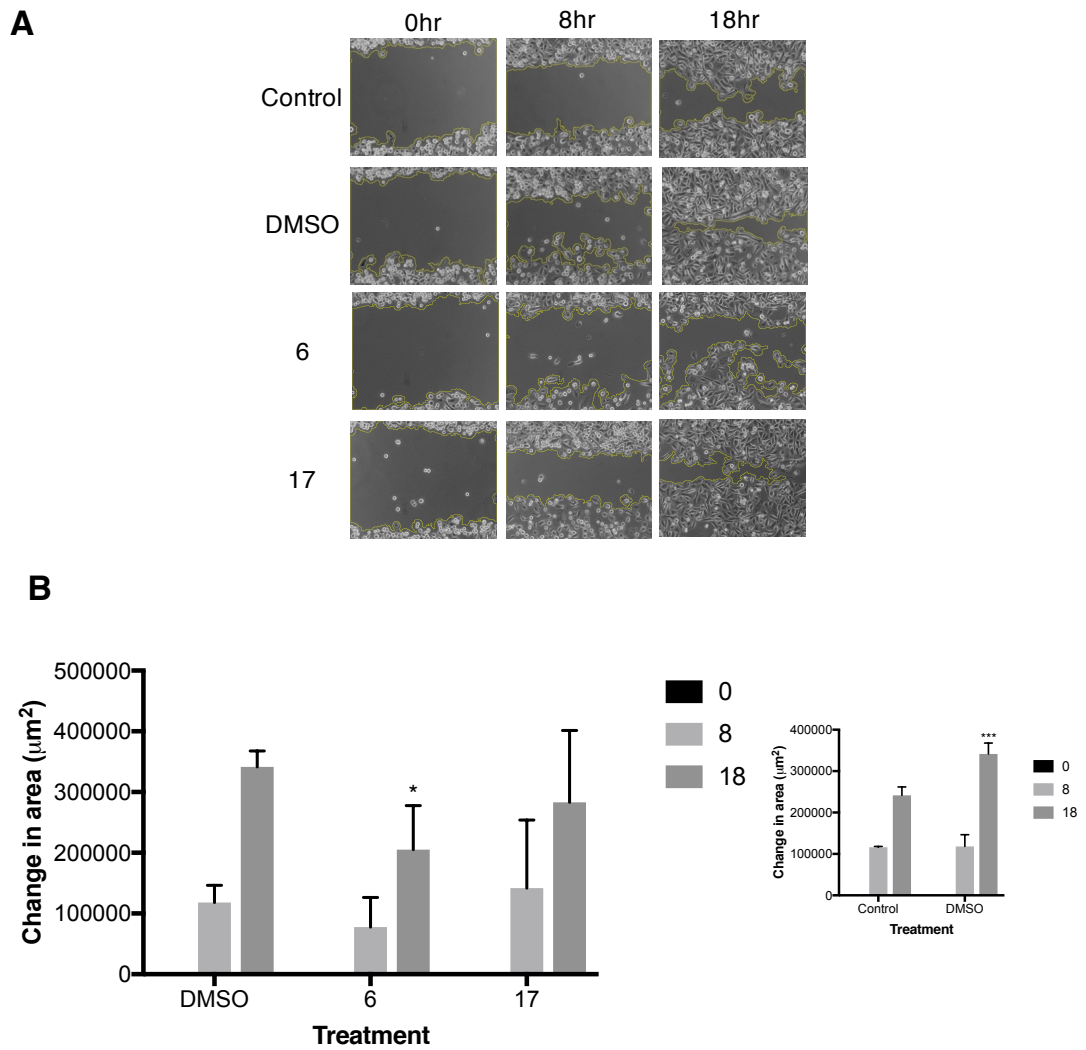


Figure 5.2.16 - Inhibition of PC-3 wound healing with natural products under TGF- β positive conditions.

At 100% confluency, PC-3 cells were incubated for 1 hour in 5 % FBS medium. Scratches were made, wells washed and 5 % FBS medium containing the appropriate conditions, described in section 2.3.4, added to wells. A) Images taken at 0, 8 and 18 hr time points. Data representative of 6 technical repeats (n=6). B) Quantification of the change in area of scratches; statistical analysis using two-way ANOVA, with Dunnett's (left panel) and Sidak's (right panel) multiple comparisons; * $P < 0.05$, *** $P = 0.001$.

Similar to the results from the scratch assay with PC-3 in TGF- β negative conditions, only compound 7 presented a significant decrease in growth/migration compared to the DMSO control (Figure 5.2.17). No significant difference was detected with or without DMSO, supporting compound 7 as an inhibitor of wound healing when there is no stimulation of the TGF- β pathway. Interestingly, when the TGF- β signalling pathway is activated no significant difference is seen for any of the compounds (Figure 5.2.18) compared to the DMSO control. This contrasts with the results of the PC-3 +TGF- β where compound 1, 7 and 19 showed a response to the presence of compound (Figure 5.2.14).

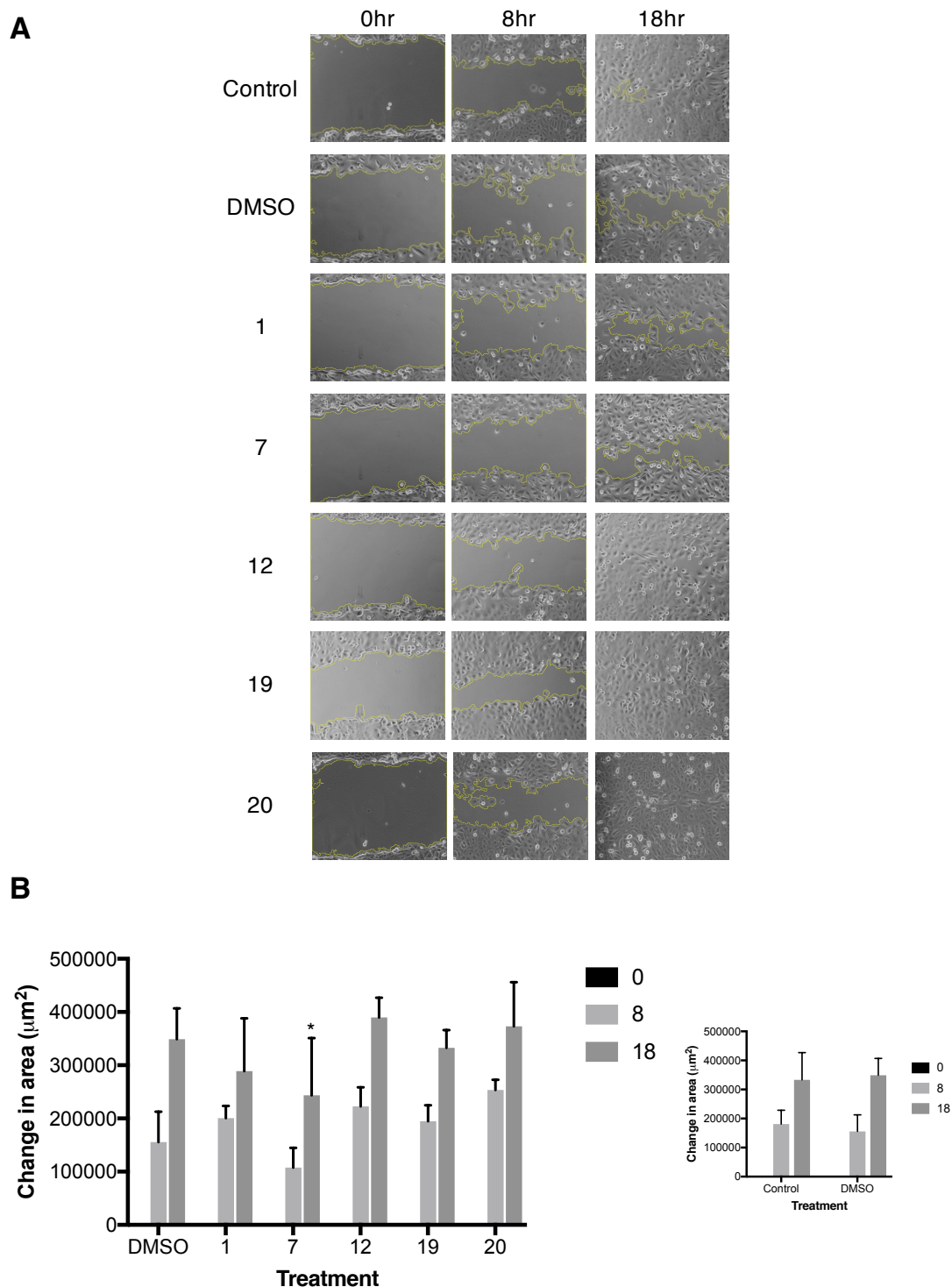


Figure 5.2.17 - DU145 scratch assays under TGF- β negative conditions.

Scratches were made in the DU145 monolayer after 1 hr incubation in 5 % FBS medium. Wells were washed and fresh media containing no supplements, 0.1 % DMSO or 10 μM compound at 0.1 % DMSO was added to the appropriate wells. A) Images taken at time points: 0, 8 and 18 hr. Data representative of 6 technical repeats ($n=6$). B) Quantification of change in scratch area at time point in A; statistical analysis using two-way ANOVA, with Dunnett's (left panel) or Sidak's (right panel) multiple comparisons, * $P < 0.05$.

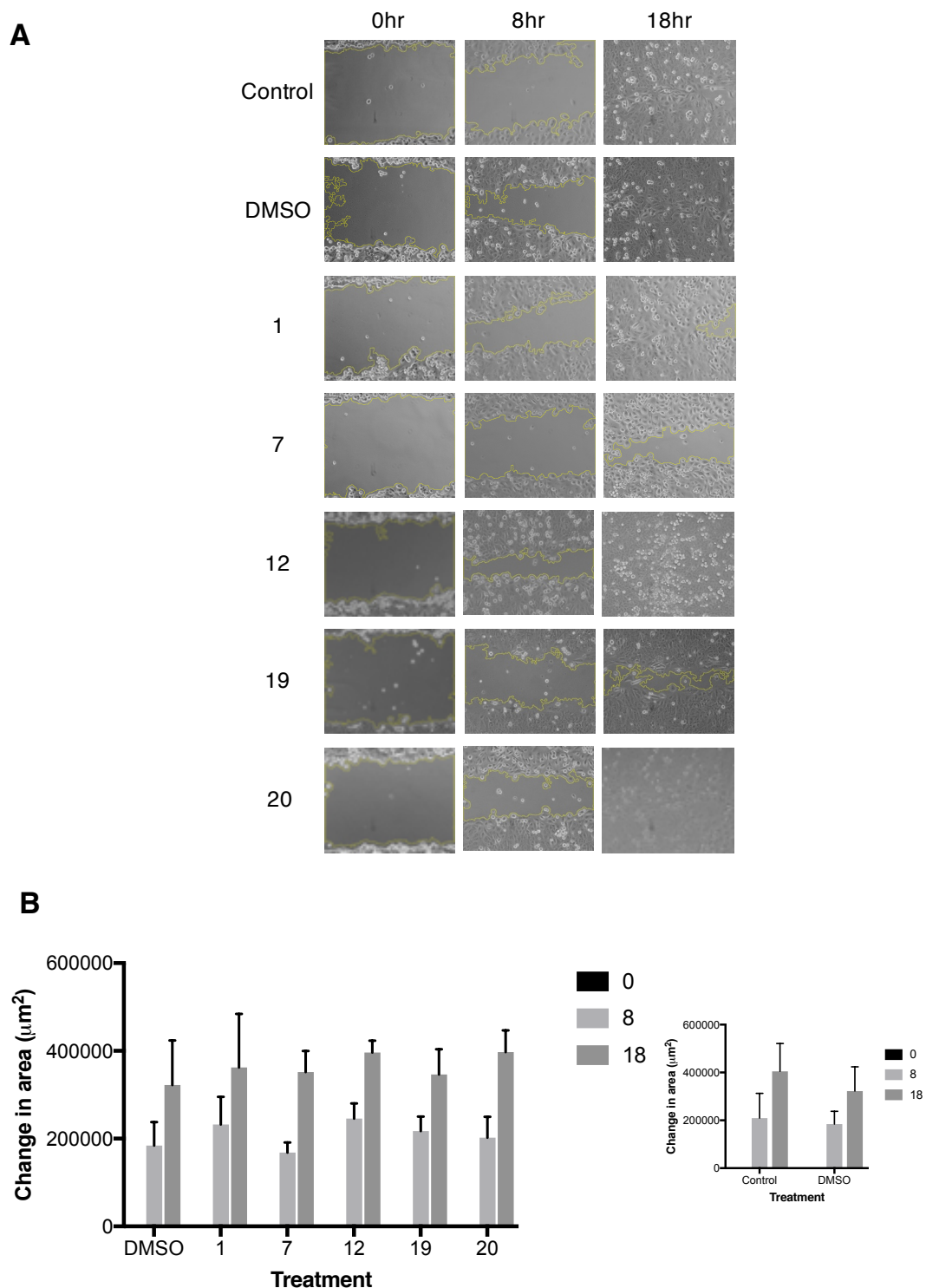


Figure 5.2.18 - Diversity set compound DU145 scratch assays in the presence of TGF- β .

After 1 hr incubation in 5 % FBS medium, scratches were made in the DU145 monolayer. Wells were washed and fresh media containing no supplements, 0.1 % DMSO or 10 μM compound at 0.1 % DMSO was added to the appropriate wells. A) Images taken at time points: 0, 8 and 18 hr. Data representative of 6 technical repeats (n=6). B) Quantification of change in scratch area at time point in A; statistical analysis using two-way ANOVA, with Dunnett's (left panel) or Sidak's (right panel) multiple comparisons, $P > 0.05$.

The results from the natural products assays show that there is no significant difference between the DMSO treatment and compounds (Figure 5.2.19), however there is a significant reduction in wound healing between the control and DMSO (Figure 5.2.19B right panel). Due to the effect shown by DMSO, comparison of the compounds with the control was carried out and showed a significant reduction in growth/migration for compound 6 but not 17 ($P < 0.01$) suggesting that the decrease in wound healing in this case is most likely due to DMSO. This differs from the results shown for PC-3 in which compound 6 produced a response. No significant difference was observed between the control and DMSO in TGF- β positive conditions (Figure 5.2.20). A significant decrease in growth/migration was observed for compound 17 but not compound 6 when stimulated with TGF- β . While this is different to the -TGF- β results, this also differs with the PC-3 +TGF- β data where it was compound 6 which showed a reduction in wound healing. This suggests a difference in the cell line responses to these compounds in the absence and presence of TGF- β , and requires further investigation.

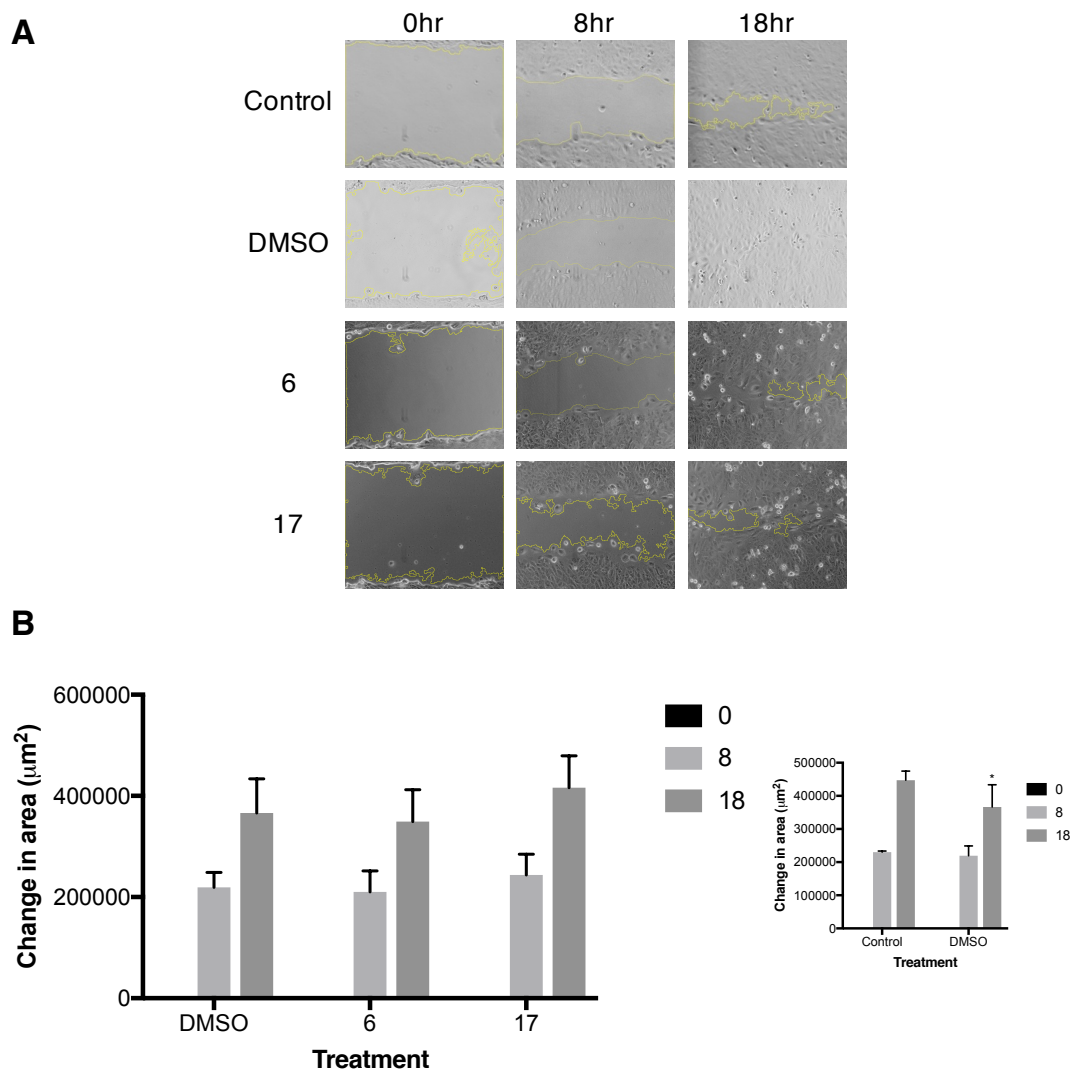


Figure 5.2.19 - DU145 scratch assay analysis of natural products 6 and 17 under TGF- β negative conditions.

In the DU145 monolayer scratches were created, and cells washed. Fresh media containing no supplements, 0.1 % DMSO or 10 μM compound at 0.1 % DMSO was added to the appropriate wells. A) 0, 8 and 18 hr time point images. Data representative of 6 technical repeats ($n=6$). B) Quantification of change in scratch area at time point in A; statistical analysis using two-way ANOVA, with Dunnett's (left panel) or Sidak's (right panel) multiple comparisons, * $P < 0.05$.

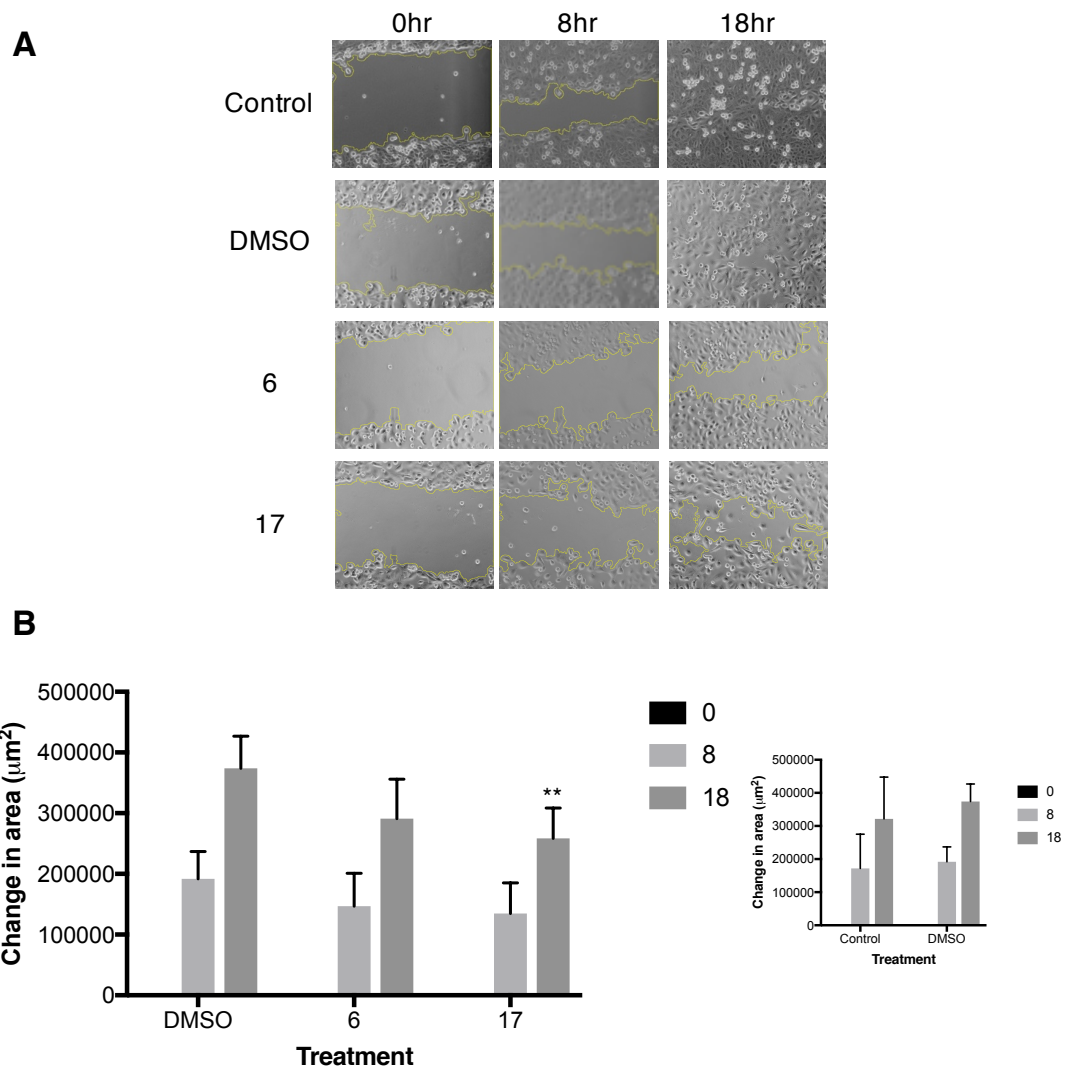


Figure 5.2.20 - Natural product DU145 scratch assay analysis in the presence of TGF- β . After 1 hr incubation in 5 % FBS medium, scratches were made in the DU145 monolayer. Wells were washed and fresh media containing no supplements, 0.1 % DMSO or 10 μM compound at 0.1 % DMSO was added to the appropriate wells. A) Images taken at time points: 0, 8 and 18 hr. Data representative of 6 technical repeats ($n=6$). B) Quantification of change in scratch area at 0, 8 and 18 hrs; statistical analysis using two-way ANOVA, with Dunnett's (left panel) or Sldak's (right panel) multiple comparisons, ** $P < 0.001$.

5.3. Discussion

Targeting the ubiquitination-proteasome system is of interest in cancer therapy. By using a method of selectively inhibiting an individual protein or group of proteins in this pathway and in turn alter their substrate targeting, this can potentially relieve the loss of protective tumour suppressor proteins. There has been progress in the development of small molecule inhibitors against the proteasome, such as Bortezomib and Carfilzomib. Bortezomib acts to block the activity of the proteasome in degrading proteins leading to a cellular build-up and cytotoxic consequences in cancerous cells (Orlowski *et al.*, 2002; Richardson *et al.*, 2003; Richardson *et al.*, 2005). While similar in activity, Carfilzomib acts in a more selective and potent manner than Bortezomib on the proteasome (Demo *et al.*, 2007). The success of these therapies however has been limited to the treatment of myeloma and lymphoma, as well as by the side effects and development of resistance. Efforts are being made to develop derivatives with higher potency, fewer off-target effects and reduce the chance of resistance (Edelmann *et al.*, 2011; Zhang and Sidhu, 2014). Consequently, attention was diverted towards the upstream proteins in the ubiquitination pathway. The discovery of RING E3 ligase inhibitors showed the potential in targeting the selective interactions as anti-cancer therapies. The overall aim of this chapter was to discover compounds with inhibitory activity against HECT E3 ligases, in particular WWP2, as potentially novel therapeutics. Initial work involved compound screening which successfully identified 28 inhibitors of WWP2-FL. Of these 28 compounds, 15 were chosen for further investigation after validation, then dose-dependent assays were performed to determine their effectiveness. Due to the ubiquitination process requiring 3 enzymes, the available compounds were also tested in counter assays against E1 and E2 enzymes. This revealed whether the inhibition detected so far was occurring in earlier stages of the pathway or directly upon E3 ligase activity. Next, the selectivity of the inhibitors was examined using two other members of the Nedd4 family, WWP1 and Nedd4. At this stage the final set of compounds to take forward was decided based on those with IC₅₀ values below 3 μ M, indicating high effectiveness towards WWP2-FL. Substrate ubiquitination using PTEN was also used to determine if these compounds affected the ability of WWP2-FL in targeting and labelling substrates or whether the inhibition was isolated to auto-ubiquitination. Lastly, in vitro analysis using PC-3 and DU145 advanced prostate cancer cell lines revealed potential inhibition of cancer cell growth/migration.

Previous work by Rossi *et al.* (2014) established a proof of concept screening method using an Itch E3 ligase construct lacking the N-terminal C2 domain, a region later found to induce auto-inhibition of the ligase (Riling *et al.*, 2015). This provided an experimental basis for the screening of compounds against WWP2, which could potentially work using truncated proteins rather than full-length. The C isoform, discussed in section 1.4.1, was of interest for this assay due to the presence of the HECT domain with the catalytic cysteine residue and the lack of C2 and WW1-3 domains, thereby reducing the possible number of ligand binding sites within WWP2. Despite showing auto-ubiquitination capabilities in initial studies (section 5.2.1), WWP2-C auto-ubiquitination activity could not be detected in 96-well plate-format experiments. In comparison, WWP2-FL produced significant auto-ubiquitination levels and the assay was optimised for use with the full-length protein. The large standard deviation shown in the optimisation results could be due to varying expression levels of GST-WWP2-FL as fresh cell lysate was used to bind the ligase directly to the plates. This method was used because of prior problems trying to re-bind purified GST-tagged proteins (data not shown). The drastically different results between the 2 isoforms of WWP2 may be solely due to a lower number of residues available for ubiquitin binding in the C-isoform. Alternatively, this may suggest that WWP2-C is unable to ubiquitinate itself but requires interaction with another WWP2-C moiety or perhaps WWP2-FL for the activation of ubiquitination, which would be limited upon immobilisation and purification on the plate. The initiation of ubiquitination activity by isoform interaction has already been discovered for WWP2, in which WWP2-N is capable of interacting and activating WWP2-FL under TGF- β negative conditions (Soond *et al.*, 2013). E6-AP has also shown increased activity upon dimerization/trimerisation of the HECT, although has yet to be shown for Nedd4 E3 ligases (Ronchi *et al.*, 2014). Another mechanism may be that activation only occurs in an adaptor protein or substrate dependent manner. The identification of Nedd4 family interacting proteins 1 and 2 (NDFIP1 and NDFIP2) showed the presence of these proteins relieved the auto-inhibitory conformations of several Nedd4 family E3 ligases, in turn activating ubiquitination as well as facilitating substrate recruitment to the ligases (Jolliffe *et al.*, 2000; Mund and Pelham, 2009; Oliver *et al.*, 2006; Trimper *et al.*, 2017). Further work by Mund and colleagues (2018) identified substrate clustering as an additional method of regulating ligase activity. It is proposed that by aggregating PY-motif-containing substrates, the activity of Nedd4 E3 ligases towards these substrates is favoured over reversion to their auto-inhibitory conformation (Mund *et al.*, 2018). The self-regulation of the individual WWP2 isoforms

is currently not clear; however, this could be another area of research in understanding the turnover of these enzymes under both normal and aberrant conditions.

After screening 2141 compounds in single-shot auto-ubiquitination assays, 15 compounds were identified as potential hits. The overall hit rate at the thresholds set in section 5.2.3 were 0.94 % and 1.2 % for the diversity set and natural products, respectively. This suggests that despite using relatively small libraries of compounds, there was still a low hit rate, implying that small molecule inhibition of WWP2 is uncommon, which is similar to the findings by Rossi *et al.* (2014) for Itch ligase. Upon further analysis, however, not all hits from the single shot assay were later identified as true. This suggests the presence of assay artefacts. These compounds most likely absorb/emit light at the wavelength used in the assay. One way of overcoming this is to initially screen the compounds to account for additional signal from compound interference. An alternative but more extensive possibility would be to alter the assay design to either to shift the wavelength of data collection to a region where fewer compounds absorb/emit energy or collect at two or more wavelengths according to the method of detection so the interference cancels out. For comparative analysis to the enzymatic screening results, other methods were also used in the lab for inhibitor screening including virtual screening (carried out by Sam Walpole) and thermal shift assays (carried out by Greg Hughes) using the HECT domain of WWP2 due to the availability of the HECT structure (PDB: 4Y07) and difficulties purifying WWP2-FL. Virtual screening showed only compound 1 at position 13/53, and the 5 natural products at positions 406/968 or lower. Furthermore, only natural product 6 was identified in thermal shift assays screened at 10 μ M compound. This highlighted the power of enzymatic assays in finding WWP2 inhibitors, and that experimental choice in screening can play a major role in small molecule inhibitor detection. Further investigations into the binding site of some of these compounds utilised ligand-based NMR techniques, as previous thermal shift analysis had failed to detect binding for all except one potential hit. The data revealed 6 of the diversity set compounds and 2 natural products interacted with the WWP2 HECT domain (carried out by Sam Walpole and Serena Monaco; modelling of the putative HECT binding site carried out by Sam Walpole shown in Figure 5.3.1). As the HECT domain is composed of 2 lobes, the C-lobe with the catalytic cysteine, and the N-lobe with E2~Ub binding site, it would be interesting to determine the precise location within this domain where these compounds interact.

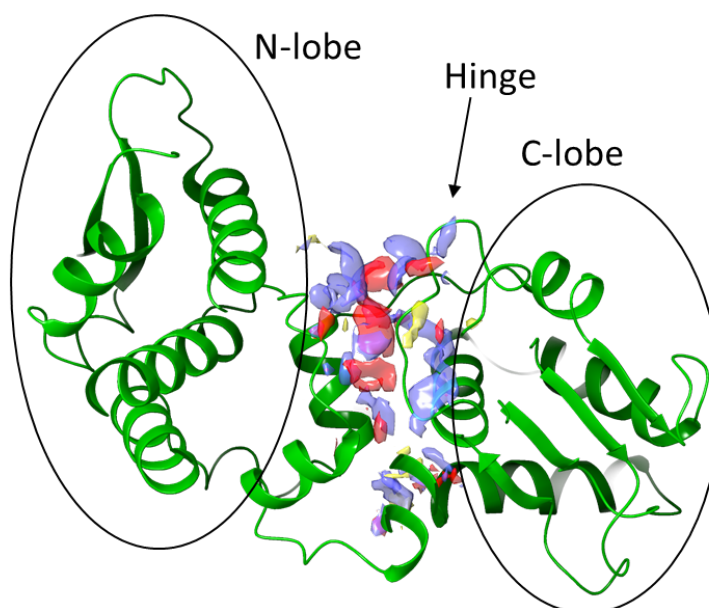


Figure 5.3.1 - Structural modelling of the putative binding site in HECT domain of WWP2.

Representative diagram of the binding site surface calculated using Sitemap and modelled by Sam Walpole (personal communication). Hydrophobic sites are yellow, hydrogen bond acceptors red and donors purple. Image generated by Sam Walpole (Watt *et al.*, 2018).

Work published by Mund *et al.* (2014) showed Heclin binding was predominantly located in the C-lobe towards the catalytic cysteine, with weaker binding in the hinge loop and N-lobe. This pattern of binding is thought to hinder the transfer of ubiquitin from the E2 enzyme to the intermediate cysteine residue by blocking the rotational motions of the lobes. It is possible that this is similarly the case with the compounds discovered so far that bind to the HECT domain, however investigations into whether ubiquitin is able to conjugate with the catalytic cysteine, or whether the inhibition may be due to the inability to re-orientate the 2 lobes remains to be answered. Preliminary data does suggest that these compounds may act upon the hinge loop on the opposing side of the HECT domain from the cysteine residue, thereby preventing the flexibility about the hinge rather than direct inhibition of the cysteine residue or E2~Ub binding site. The HECT domains of Nedd4, Nedd4-2 and Smurf 2 were shown to contain a non-covalent ubiquitin binding site, proposed to be involved in poly-ubiquitination (Maspero *et al.*, 2011), providing another potential binding site for inhibitory activity by small molecules. Kathman *et al.* (2015) showed inhibition of the non-catalytic cysteine present in Nedd4 that is involved in poly-ubiquitin chain formation by small molecules, supporting this as a mechanism of inhibition for member of the Nedd4 family. However, as WWP2 appeared to lack such ubiquitin-binding capabilities of HECT in the study,

and a cysteine residue in the equivalent position, this suggests that this site may not be present in WWP2 and another mechanism may be employed. However, WWP2, WWP1 and Itch do share a cysteine residue further upstream which may be involved and could be an equivalent site for facilitating poly-ubiquitination and in turn inhibition, though this has yet to be investigated. 3 compounds showed no interaction with the HECT domain but showed inhibition of WWP2-FL activity, suggesting that these compounds bind outside of the HECT or bind with affinities too strong to be detected by ligand-based NMR. This then raises questions about how these compounds are inhibiting auto-ubiquitination if they bind outside the HECT domain – are they inducing an inhibitory conformation of the ligase? This may be possible via interaction with WW2-3 linker region that interacts with the HECT domain and auto-inhibits WWP2 activity (Chen *et al.*, 2017), or is there some other allosteric regulatory mechanism induced by the small molecule binding that is currently unknown? Alternatively, if these compounds are binding extremely tightly to HECT, this potentially suggests that these may be irreversible inhibitors of WWP2 with slow dissociation from the enzyme. These questions however remain to be investigated. Difficulties regarding stability of the remaining 3 compounds under the ligand-based NMR conditions meant that binding was not determined. Collectively, the combination of the enzymatic assay and ligand-based NMR supports the identification of compounds capable of inhibiting WWP2 by interaction with the HECT domain and opens up new avenues of potential mechanisms of inhibition.

In order to confirm the inhibition of WWP2 and conclude that the point of inhibition is in the last step of ubiquitination, further analysis was carried out to elucidate any potential inhibitory action on E1-activating and E2-conjugating enzymes upstream in the ubiquitination sequential cascade. Uba1 is a human E1 activating enzyme which can interact with many E2-conjugating enzymes. In a similar manner, E2-conjugating enzymes can interact with many E3 ligases, therefore it is key to not alter the function of these proteins significantly. Xu *et al.* (2010) showed that E1 inhibition by siRNA and small molecules could induce malignant cell death, therefore investigating whether these compounds have an effect of E1 or even E2 may also open up new avenues of research. The results of the E2 counter screen showed no evidence for inhibition of the E1 or E2 activity. However, E1 only assay results suggested otherwise. Upon analysis of E1 ubiquitination using an alternative method in order to visualise E1~ub complex formation, this highlighted that there was no complete inhibition of E1 activity, hence why no significant effect was observed in the E2 assays. This may suggest an assay

development problem regarding Uba1 inhibition analysis, however, investigation of the E1 enzyme inhibitor PYR-41 showed partial inhibition of HECT E3 enzymes, suggesting that perhaps these compounds display similar inhibition mechanisms to PYR-41 (Yang *et al.*, 2007). This supports the results of the plate assay being a true reflection of partial inhibition of E1 but with no significant effect on the next step of the ubiquitination cascade. While these results are not ideal for the discovery of selective WWP2 inhibitors, it opens up a potential area of future investigation into possible novel E1 inhibitors, which remains a relatively unexplored field for drug discovery in cancer therapeutics (Itoh and Suzuki, 2018).

Another aspect analysed was the selectivity of the 14 compounds. 4 compounds showed significant inhibition of all 3 Nedd4 E3 ligases, with no significant difference in % residual activity levels between the 3 proteins. This potentially indicates that these compounds bind and inhibit these ligases in a very similar manner. 3 of these 4 small molecules have shown HECT binding in ligand-binding analysis, mentioned previously, but one was inconclusive. This supports a mechanism of interaction that is shared between the HECT domains of the ligases, which considering the similarities between HECT domain structures and motions would not be surprising. In contrast with this, 5 out of 14 hits had a greater effect on WWP2 inhibition in comparison to Nedd4 and WWP1. 4 of these also happened to interact with the HECT domain, as mentioned above. This suggests these compounds may be manipulating the differences observed between the HECT domains or may just have a better fit or increased affinity when interacting with WWP2 compared to WWP1 and Nedd4. Studies into the structures and movements of the HECT domains of WWP2, WWP1 and Nedd4 have shown differences in the rotational ability and binding sites (Gong *et al.*, 2015; Maspero *et al.*, 2011; Verdecia *et al.*, 2003). Interestingly, the intramolecular interactions between the N- and C-lobes of WWP2 and WWP1 have been found to differ (Gong *et al.*, 2015), which may be one of the ways which these compounds are showing increased inhibition towards WWP2 over WWP1 despite these proteins being very closely related. In contrast, natural product 6 showed greater inhibition of Nedd4 than WWP1 and WWP2 while also showing the ability to interact with WWP2 HECT, highlighting that the interaction with Nedd4 may be more favourable and could present a possible inhibitor for development towards Nedd4 over WWP2. As for the remaining diversity set compounds that were tested, both presented increased inhibition of WWP2 and Nedd4 compared to WWP1 with one able to bind WWP2 HECT and the other inconclusive. It would be interesting to determine whether the binding site resides

outside of the HECT domain, as Nedd4 and WWP2 have been shown to share substrates which could reveal a potential inhibitor of substrate interactions over inhibition of HECT domain activity. Natural product 15 produced similar results with even lower inhibition of WWP1 than the other compounds produced, and has not been found to bind WWP2 HECT. Lastly, natural product 17 showed inhibition of WWP2 and WWP1 to a greater degree than Nedd4, suggesting that this compound manipulates the close similarity between these two ligases compared to Nedd4. In order to say more conclusively whether these differences between the E3 ligases are due to possible alternative binding sites or affinities, more experimental data is required into the binding of these compounds with WWP1 and Nedd4.

Auto-ubiquitination self-regulates the activity of HECT E3 ligases and is important for the control of substrate concentrations at all times. WWP2 has been shown to interact with tumour suppressor protein PTEN (Chen *et al.*, 2016; Maddika *et al.*, 2011), which has been implicated in prostate cancer development and progression. Substrate ubiquitination using PTEN in the presence of potential inhibitors showed 6 compounds prevented poly-ubiquitination of PTEN, however compound 6 still showed some generation of high molecular weight PTEN~Ub complexes. This suggests compound 6 infers greater inhibition on auto-ubiquitination of WWP2, than on PTEN ubiquitination, although this would need to be looked at in more detail. It would be interesting to see if similar inhibition is observed for ubiquitination of PTEN by Nedd4, as PTEN is a known substrate of Nedd4 as well, although WWP2 is the preferred ligase when PTEN is unmodified (Chen *et al.*, 2016). In combination with the results of the selectivity assays, compound 6 shows potential for inhibiting Nedd4 over WWP2 in which case could also prevent the degradation of PTEN when it is modified. By reducing the ability of WWP2 to target PTEN for degradation, the resultant increase in PTEN could negatively feedback in the AKT signalling pathway leading to a decrease in cellular survival and growth. This has been observed using siRNA against WWP2 and a WWP2^{C838A} mutant by Maddika *et al.* (2011), stabilising PTEN levels and reducing the oncogenic potential of cancer cells. Whether this inhibition of substrate ubiquitination is mirrored by other proteins targeted by WWP2 remains unanswered and should be investigated later.

The wider impact of this research relates to developing novel inhibitors that are effective as a treatment for prostate cancer. PC-3 and DU145 cell lines were chosen as they lack an androgen response but retain TGF- β sensitivity (Cunningham and You, 2015;

Kang *et al.*, 2001; Wu *et al.*, 2013). This ensures that any changes observed are not due to the influence of the TGF- β signalling pathway on the AR pathway (Kang *et al.*, 2001). Furthermore, PTEN expression is absent in PC-3 cells but present in DU145 (Lee *et al.*, 2004). These factors are important in this research as both PTEN and Smad proteins in TGF- β signalling are target substrates of WWP2. Using these 2 cell lines provides evidence of compound effects under different molecular backgrounds of prostate cancer, however ideally these compounds would be tested in primary tumour samples to account for the heterogeneity of prostate cancer (Ciccarese *et al.*, 2017; Cunningham and You, 2015). In the absence of TGF- β stimulation, both cell lines showed a reduction in migration in the presence of diversity set compound 7, but that is where the similarities in response end. PC-3 migration was also reduced by natural product 6. Under TGF- β stimulation, only natural product 17 produced a decrease in migration of DU145 cells. However, PC-3 cells were slowed by compounds 1, 7, 19 and natural product 6. Without evidence of the protein and mRNA levels of key components in the TGF- β pathway, PTEN and WWP2, and evidence of compound uptake, it is difficult to make any conclusions about what might be happening in these cell lines on a biochemical viewpoint. However, there is evidence that some of these compounds show potential anti-oncogenic effects for advanced prostate cancer.

An important aspect to take into consideration when using biochemical screens for small molecular inhibitors is the potential for false positives. This may be caused by the aggregation of compounds leading to non-specific inhibition or by pan assay interference compounds (PAINS) (Gilberg *et al.*, 2016). Research has gone into the generation of computational filters to try and identify potential PAINS in compound libraries due to high number of compounds with promiscuous binding properties. Analysis of the 5 hits generated in this work from the NCI diversity set V highlighted compound 7 (228155) as a known PAIN molecule (Jasial *et al.*, 2017), which could also indicate compound 19 (228150) as a potential PAIN molecule and should be taken into consideration in further investigations. Furthermore, the quinone groups that make up compound 1 (2805) are well-known PAINS therefore 2805 may also be a false-positive to be carefully considered in further work. As for the remaining diversity set compounds, 12 (44750) and 20 (288387), when filtered for potential PAINS, no hits were produced, suggesting these are not false-positives and are most likely compounds with true activity within these assays. In a similar search on the 2 natural products, no hits were produced for natural product 6 (401005), and only 1 hit was generated for natural product 17 (169517), in which a derivative called quercetin containing the PAINS

catechol group was used in the research (Baell and Holloway, 2010). While natural product 17 lacks this catechol PAINS motif, in its place exists a quinone group which is also present within natural product 6. This suggests that these compounds could potentially have some PAINS-based redox activity, however, as these have not specifically been identified as interference compounds, may yet provide a basis for further investigations but should be done so with caution. Overall, PAINS analysis revealed 2 potential hits that produced no PAINS hits, which in conjunction with having low μM IC₅₀ values against WWP2, supports these compounds as potential hits for further analysis.

6. Discussion

6.1. Introduction

WWP2 is a Nedd4 family HECT E3 ligase that targets substrates with PPxY motifs for ubiquitination. Well known substrates of WWP2 include the tumour suppressor proteins PTEN and Smad 2, 3 and 7. The overexpression of this E3 ligase has been linked to prostate cancer development and progression via the increased degradation of these substrates. Therefore, the aims of this work were to build upon current knowledge to facilitate the discovery of novel inhibitors against WWP2. Two main approaches were used: NMR spectroscopy was carried out to determine the structure of the fourth WW domain of WWP2 to enable the design of potential peptidomimetics (Results chapter 1); and titration analysis was utilised to understand the substrate interactions of the first and second WW domains in the hopes of using this to determine the subtle characteristics that may further direct peptide design to selectively inhibit substrate interaction (Results chapter 2); secondly, compound screening and biochemical analysis of potential small molecule inhibitors against WWP2-FL enabled the investigation of an activity-based approach with compound inhibitors (Results chapter 3). Combined, these approaches can direct selective drug design towards this ligase, expanding upon the few that are currently published against other members of the Nedd4 family, in order to advance towards a novel therapeutic in the future.

6.2. Structure-based inhibitor discovery

Structural solution of WW4 in WWP2, discussed in chapter 3, was performed using previous data by Lloyd Wahl (2016) and additional experimentation in this research with a double-labelled GB1-tagged WW4 construct. The final ensemble consisted of a 3 stranded β -sheet for WW4, reflective of the common folding pattern of HECT ligase WW domains solved to date. Out of the 24 WW domains of the Nedd4 family that have currently been structurally solved, the one with the most similarity by RMSD was WW2 of WWP2 itself (PDB: 5TJ7 – WW2-2,3-linker-HECT; Chen *et al.*, 2017) at a value of 0.824 Å, calculated using Chimera structure comparison of medoid model 7 for WW4 and chain B WW2 region. Interestingly, the next most similar WW domain in the Nedd4 family in the PDB is HECW1 WW2 at 1.043 Å (PDB: 3L4H; Walker *et al.*, 2010). Comparison of sequence similarity highlighted that WW2 of HECW1 is actually more similar to WW4 than WWP2 WW2 (Figure 6.2.1A). This supports the idea that WW domains within the same protein have unique roles to play when it comes to substrate selection. Furthermore, the similarity in the primary structure of these WW domains in their unbound conformations suggests that their substrate selection preferences may occur via subtle changes in specific amino acids. This is likewise supported by

preliminary data that the WW domains of WWP2 actually have the ability to bind to the same substrates, for example both WW2 and WW4 can bind to Smad 7 substrate, and similarly WW1 and WW4 have been shown to interact with Smad 2 and Smad 3 (Danielle Bourcier, 2018, personal communication; Wahl, 2016; Soond *et al.*, 2013). With regards to the three solved structures mentioned above, there are, however, some noticeable differences. Both the first and second β -strands are shorter by one amino acid in WWP2 WW2 resulting in a longer loop 1 region (Figure 6.2.1B). This variability which is also present when compared to HECW1 WW2 may support the reduced role this region plays in WW domain function also reflected in the amino acid sequence variability of this region.

A

WWP2 WW2	330	RPL	PPG	WEK	RTD	PRGRF	YYVD	HNTR	TT	WQR	PTA	363	(34)
WWP2 WW4	444	PAL	PPG	WEM	KYT	SEGV	RYFVD	HNTR	TT	FKD	PRP	477	(34)
HECW1 WW2	1018	LEL	PRG	WEI	KTD	QQGKS	FFVD	HNSR	AT	T	FID	PRI	1051 (34)
		**	***	:		.	*	:	*****	:	***	:	*

B

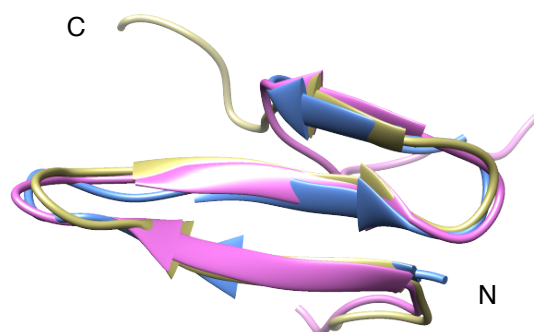


Figure 6.2.1 - Structural comparison of WWP2 WW2, WW4 and HECW1 WW2 domains.

A) Sequence comparison of the WWP2 WW2 and HECW1 WW2 with WWP2 WW4. * represents 100 % similarity; : means strongly similar in properties; . refers to weak similarity in properties. Green – conserved tryptophan residues, or atypical phenylalanine residue. Blue – conserved prolines in the hydrophobic core. Red – residues linked to WW domain class and substrate binding. B) Overlaid structures of WWP2 WW2 in blue (PDB: 5TJ7, chain B residues 335 - 362; Chen *et al.*, 2017), WWP2 WW4 in brown and HECW1 WW2 in pink (PDB: 3L4H, residues 1018 – 1056; Walker *et al.*, 2010).

In contrast, however, the residues of the XP pocket in WW4, which are important for the binding of the X-P region of the PPxY motif, show both similarity and divergence in

positioning compared to WWP2 WW2 and HECW1 WW2. Comparison of the 3rd strand highlighted that in the WW4 medoid model 7, the rotation of phenylalanine was slightly more parallel to tyrosine unlike the WW2 domains of WWP2 and HECW1 which were slightly more towards a perpendicular orientation (Figure 6.2.2A). Phenylalanine in WW4 models 16 and 20, in which the XP pocket residue is defined as β -strand in nature, shows closer similarity in side chain position to the WW2 domains (Figure 6.2.2B). This could either be the result of a reduced number of distance restraints present for the aromatic residue or could be an indication of conformational sampling that might be linked to substrate selection. The latter is supported by the research of Panwalker and colleagues (2016) who found that with Nedd4 WW3 the 3rd strand underwent the most changes in stability and rigidity upon substrate interaction. The comparison of published structures of bound conformations supports the potential that WW4 models 16 and 20 present a deviation towards that of bound conformations. For example, Smurf 1 WW2 second tryptophan when interacting with Smad 7 peptide (PDB: 2LTX; Aragon *et al.*, 2012) mirrors the more perpendicular orientation similar to models 16 and 20 for WWP2 WW4 (Figure 6.2.2C). Likewise, Smurf 2 WW3 bound to Smad 7 peptide (PDB: 2KXQ; Chong *et al.*, 2010) has a more open conformation in a more perpendicular fashion (Figure 6.2.2D), however is much further away from the tyrosine than WW4 is, suggesting that the rotational flexibility of this residue is important for the hydrophobic interaction with proline. It would be interesting to determine the XP pocket orientation of WW4 when bound by Smad 7 for comparison to determine whether this residue also orientates into the more open conformation upon substrate binding. Further research into the dynamics of this region in both unbound and bound configurations is required to understand this in detail.

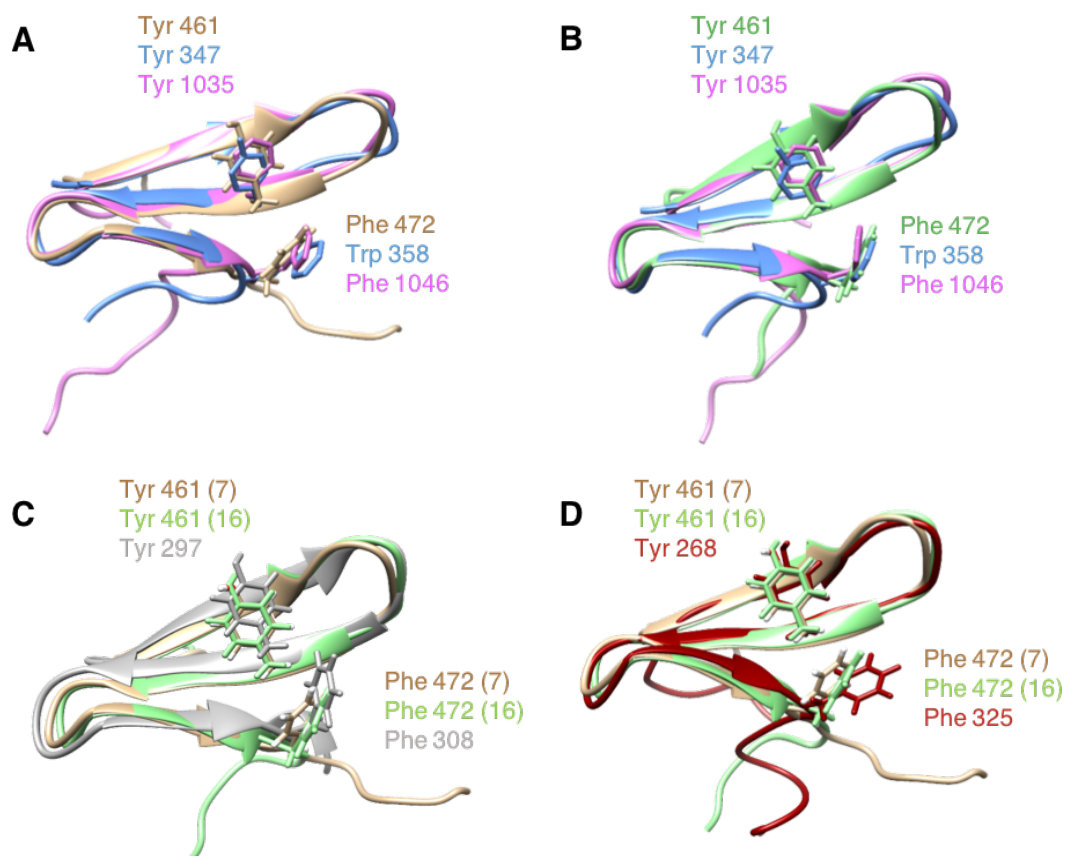


Figure 6.2.2 - XP pocket comparison of WWP2 WW4 models with the second WW domains of WWP2 and HECW1.

A) Overlaid structures of medoid models for WWP2 WW4 (brown), WWP2 WW2 (blue) and HECW1 WW2 (pink). B) Structure comparison of WWP2 model 16 (green) with WWP2 WW2 (blue) and HECW1 WW2 (pink). C) Comparison of WWP2 WW4 medoid model 7 (brown) with model 16 (green) and Smad 7 bound Smurf 1 WW2 (grey). D) Comparison of WWP2 structures as described in C with Smad 7 bound Smurf 2 WW3 (red). All structures show the side chain of the XP residues – tyrosine on the 2nd strand and phenylalanine/tryptophan at the end of the 3rd strand.

Interestingly, when looking at the other aromatic residue of the XP pocket, tyrosine 461 (phenylalanine in HECW1 WW2) has a much better-defined orientation across the ensemble as well when compared to the same position in WWP2 WW2 and HECW1 WW2. This could be due to the central nature of the residue in the structure in turn having more surrounding connections and a more restrained side chain orientation in structural calculations. Alternatively, this may suggest that the orientation of this residue in this apparent fixed position is vital for correct binding of substrates. Similarly, valine 463 which acts as a secondary specificity site for class I substrate interactions (Espanal and Sudol, 1999) is also tightly restrained across the WW4 ensemble (Figure 6.2.3A)

and is consistent with HECW1 WW2 with a little variation to WWP2 WW2 (Figure 6.2.3B). Unlike valine 463, at the + 2 position to this residue the histidine (H465) previously thought to be involved in substrate specificity has a reduced involvement despite being conserved (Kasanov *et al.*, 2001; Wahl, 2016). This is supported within the WW4 ensemble where H465 presented increased diversity in the side-chain orientations (Figure 6.2.3A), which is also mirrored when compared to WW2 domains in WWP2 and HECW1 (Figure 6.2.3B).

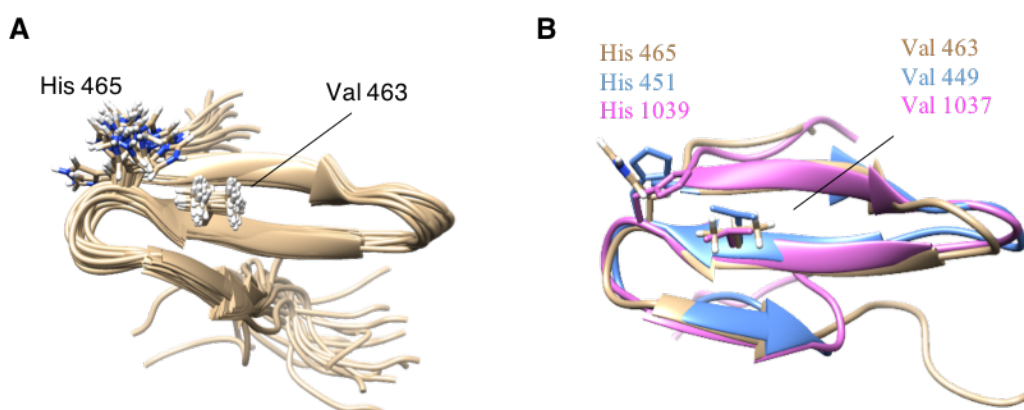


Figure 6.2.3 - Comparison of residues outside of the XP pocket involved in substrate interactions.

A) Valine 463 and histidine 465 across the refined WW4 ensemble. B) Comparison of valine 463 and histidine 465 in WW4 medoid model 7 (brown) with the equivalent residues in WWP2 WW2 (blue) and HECW1 WW2 (pink).

The N and C termini of WW4 both fold underneath the β -sheet away from the curved binding site of the domain, stabilising the conformation via hydrophobic interactions. Due to the lack of structure for the N-terminal of WWP2 WW2 before the tryptophan residue, comparison of the stabilising proline in this region could not be inspected (Figure 6.2.4A). Similarly, the ill-defined region of the WW4 domain at the C-terminal also prevented confident analysis of these residues therefore little can be said for the role of these residues in stabilising the domain other than seeming to come closer in the opposing direction of the well-defined core (Figure 6.2.4A). However, the first tryptophan at position 450 in WW4 and 336 in WW2 within WWP2 have very different side chain positions (Figure 6.2.4B). Within WW4, tryptophan seems to favour a direction towards the C-terminal tail of the construct underneath loop II with contacts to aspartic acid 464 and asparagine 466, and potentially proline 447 in the N-terminal region. In contrast, the equivalent tryptophan in WW2 appears to orientate the other way. This may be because the C-terminal tail of WW2 seems to curl under towards the

second loop of the structure with arginine 360 extending towards tryptophan 336, unlike the WW4 tail which follows a more perpendicular direction away from the β -sheet, however with no N-terminal structure it cannot be determined whether contacts here may also be involved in this difference. Similar to WW4, tryptophan 1024 in HECW1 has the same orientation as WWP2 WW4, not WWP2 WW2 but does display the same C-terminal tail positioning towards loop II (Figure 6.2.4). However, as to whether there is a significant difference in stability between these domains is unknown therefore this difference in positioning would require further investigation to determine any role this plays in functionality. One other aspect to note regarding WW2 and WW4 of WWP2 is the lack of a defined linker between the first and second domains of WWP2 unlike between WW3 and WW4 which may affect the positioning of this residue, however again this would need further analysis to understand whether this is important or not within the full-length protein.

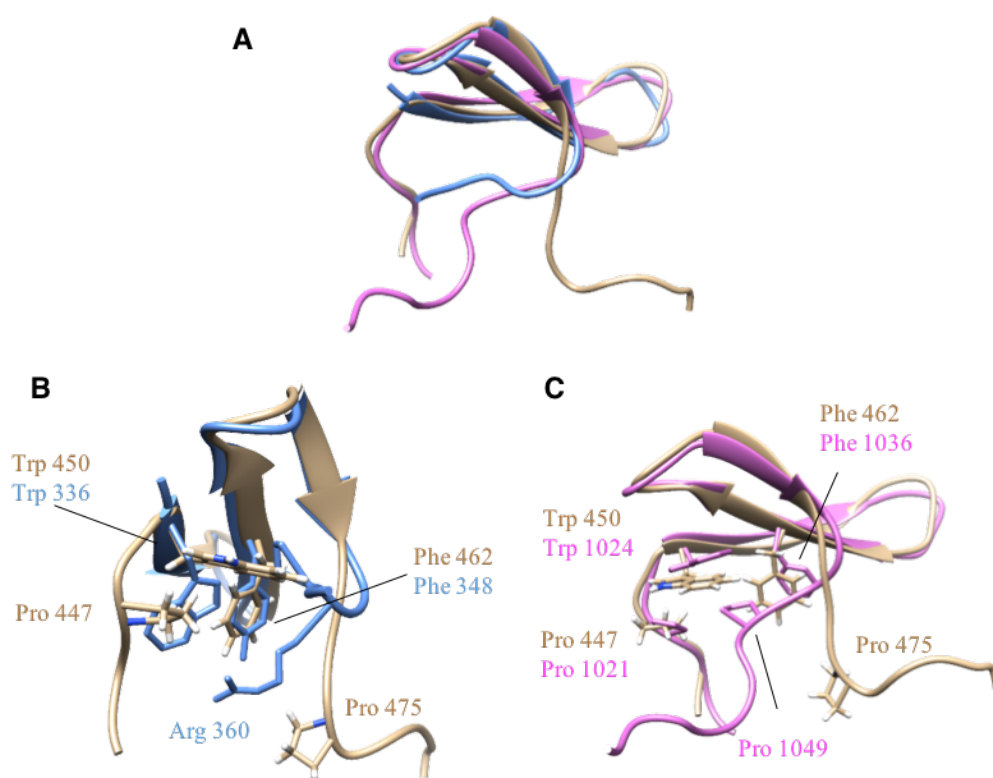


Figure 6.2.5 - Comparison of the hydrophobic underside of WWP2 WW4 with WWP2 WW2 and HECW1 WW2.

A) Alignment of WWP2 WW4 (brown) with WWP2 WW2 (blue) and HECW1 WW2 (pink) with focus on the underside of the domains. B) Comparison of the hydrophobic residues of WWP2 WW4 (brown) and WW2 (blue). C) Overlaid WWP2 WW4 (brown) and HECW1 WW2 (pink) showing the hydrophobic residues.

The biological function of this structure can be seen in the selective number of known substrates that interact with WW4. Interestingly, there was very little difference in the binding site residues involved in Smad 2, 3, and 7 interactions with WW4 (Wahl, 2016). The residues with the greatest change in environment were threonine 469, 470 and 471 located to the 3rd β -strand for all three substrates, which supports the evidence by Panwalker *et al.* (2017) with Nedd4 WW3 that this region of the WW domain may be the most affected by the binding of a substrate. However, further work is necessary to elucidate the bound structures of this domain with Smad 2, 3 and 7 to identify significant conformational changes that may be specific for each substrate and underpin the difference in affinities reported by Lloyd Wahl.

Using the conserved residues and those found to be involved in binding (Kasanov *et al.*, 2001; Wahl, 2016), the basis for initial peptide design that reflects WW4 would need to include:

xxLPxGWEMK/RxbxxGxxY/FY/FVxHxTR/KTTTFxxPxx

where x represents any aliphatic amino acid and b refers to a polar residue. As yet only the isolated WW domains of WWP2 consisting of the naturally occurring primary sequence have been analysed in vitro (Yim, 2017). Yim's preliminary research looked at the effect of individual WW domains, as well as tandem domain constructs, on Smad 3-dependent gene expression. A model by Soond *et al.* (2013) suggested that cancer initiation occurred via a reduction in Smad 3 by activation of WWP2-FL by the shorter isoform WWP2-N under TGF- β negative conditions. When TGF- β is then present, as is seen in cancerous cells which appear have a higher basal level of TGF- β than non-cancerous cells, an increase in WWP2-FL and WWP2-C containing the WW4 domain then changes substrate preference towards Smad 7 over Smad 3. Considering recent titration data, the preference of these isoforms under stimulated conditions would most likely have been towards phospho-Smad 7 which has a higher affinity to WW4 than unphosphorylated Smad 7 (Wahl, 2016). Using this model as a basis, Tiffany Yim (2017) found that expression of WW4 led to no significant difference in unstimulated control cells as well as melanoma cell lines. However, when in the presence of TGF- β stimulation, there was an increase in Smad 3-dependent gene activity in both the control and cancerous cell lines. This increase in Smad 3-dependent activity of the pathway may be advantageous in therapeutic terms for the cancer initiation stage by stabilising Smad 3 levels, but further investigations into cancer cell migration,

proliferation and the biochemical cellular changes upon the addition of this domain is required in both non-cancerous cells and at different stages of cancer progression. Interestingly, analysis of the effect of tandem domains under the same conditions revealed that WW3-4 expression increased Smad 3-dependent activity regardless of the stimulation status and cell type. It was thought that this might be due to the presence of the 3-4 linker region rather than the domains themselves. Addition of WW3-4 is hypothesised to have resulted in sequestration of Smad 3 via interaction with WW3-4, and/or potential dimer formation with the same region in WWP2-FL, although evidence of WWP2 WW domain dimerization has yet to be uncovered. Furthermore, the identification of the 2-3 linker region as a regulator of the auto-inhibition conformation of WWP2-FL (Chen *et al.*, 2017) further suggests regulatory roles for these regions and that creating a peptide based only on WW4, or even the other domains individually, may not be effective in inhibiting the activity of WWP2 but may in fact exacerbate the cancerous profile via targeting of regulatory Smad 3 and inhibitory Smad 7 in certain conditions. Therefore, further understanding of this switch between substrate choice by WW4 is required. Research into the effect these WW domain constructs have on Smad 2 and other substrates such as PTEN and Oct4 when expressed in mammalian cells should also be investigated to determine if changes in expression or signalling activity for these substrates are also affected.

Knowledge of the bound structure of WW4 to compare to the unbound reported in this thesis and using this for peptidomimetic modelling could reveal novel information about substrate selection and further direct successful design of inhibitors for this domain. Additionally, the identification of further novel isoforms that contain the WW4 region (Yim, 2017) increases the complexity of targeting this domain, especially as knowledge of the function of these new isoforms is currently lacking. Also, it is important to note that these experiments have only been carried out using melanoma cell lines so far, and the same results may not occur in prostate cancer cells, therefore it would be interesting to determine the effect of the introduction of these isolated domains to prostate cancer cell lines. Also, the design of sequence variations based on these domains and linker regions using structural modelling, similar to the medicinal chemistry approach to modifications of small compounds, might also be worth investigation in the future when more is understood about these domains and their substrates.

Another aspect to take into consideration is the ability for multiple domains within WWP2 to bind to the same substrate, as mentioned earlier. Previously, Soond and Chantry (2011) concluded that WW1 within both WWP2-FL and WWP2-N isoforms is the main interaction site for Smad 2 and 3, whereas WW4 within WWP2-FL and WWP2-C preferred Smad 7. As the interactions of the fourth domain had already been investigated, showing capabilities of WW4 in binding to all three Smad proteins of the canonical TGF- β pathway, attention was turned to first WW domain and its role in WWP2 interactions, especially since WWP2-N was found to be significantly altered in prostate cancer. Furthermore, understanding the structure and binding elements of this domain could enable multi-domain targeting by novel therapies with WW4 to provide a potential combinatorial effect, although as to whether this is effective has yet to be tested. As discussed in Chapter 4, structural solution of the first domain was not deemed possible due to the unstructured nature of the construct used in this work. Jiang *et al.* (2015) reported having problems expressing this domain for structural characterisation experiments using a GST tag, but the use of the GB1 tag in this work seemed to increase the levels enough to obtain adequate concentrations for NMR spectroscopy. HSQC analysis and backbone assignment identified 30 % of WW1 as unassignable, not including proline residues, due to missing peaks. There can be several reasons for such a high level of missing peaks in a HSQC spectrum. Firstly, it could simply be that the peaks are overlapping in the HSQC which 3D experiments would reveal, but only a few extra peaks were present in HNCaCb and CbCaCONH that were at chemical shifts of already assigned residue types and these did not account for all the missing peaks. Alternatively, these regions of the protein could be in intermediate chemical exchange, where the rate at which the population changes between different environments is proportional to the NMR time scale, resulting in broad peaks as the rate of exchange is close to the frequency separation of the environments. One way to overcome this is to alter the temperature used for the NMR experiment to either speed up (rate > chemical shift difference) or slow down (rate < chemical shift difference) the exchange rate. This can be done in either direction dependent on what is wanted from the experiment – an increase in temperature would shift the exchange rate towards fast exchange generating an average peak in the spectrum, or a decrease in the temperature can be used to shift towards slow exchange in which two peaks are present representative of the two chemical environments. Previous work in the lab (Danielle Bourcier, 2018, personal communication) using WWP2 WW2 domain, where HSQCs were collected over a range of temperatures, indicated no significant change in peak number upon a decrease in temperature from

298 K, although temperatures higher than this were not tested due to the risk of denaturing or degrading the sample. Another aspect could be the buffer or pH of the sample. Due to the success of NMR analysis with WWP2 WW4, the same sodium phosphate buffer was used for the WW1 samples at pH 6.8. Perhaps further screening of a range of buffer types and pHs often used in NMR experiments is worth investigating to determine whether there is a more compatible buffer for WW1 and in turn perhaps WW2 which also displayed similar problems. Another possibility is that this domain is not folded when isolated in this construct or even outside of the cellular environment therefore structural determination would not be possible.

The options above mainly focus on the practical aspects of the sample generation and NMR experiment conditions. Other reasons which can cause missing peaks include dimer formation in which residues of the interface between the two molecules can undergo significant changes in environments resulting in intermediate exchange. Evidence of WW domain dimerization for Smurf 1 and 2 was reported by Aragon *et al.* (2012), and in conjunction with data supporting interaction between WWP2-FL and WWP2-N (Soond and Chantry, 2011a), this suggests that dimerization between WW1 molecules may be possible. Alternatively, WW1 in its apo state may adopt an unfolded flexible conformation that only completely folds upon interaction with substrates. Collectively, this suggests that there are even more questions about WW1 that have yet to be answered and further work is essential in understanding this domain.

Investigations turned towards substrate interactions and determining the binding site residues to see whether this could provide insight into the structure and function of the domain, and whether bound conformations could be possible in the future. Unfortunately, incomplete saturation of Smad 2 peptide and protein denaturation from the presence of high concentrations of TFA in the Smad 3 titration prevented the identification of the binding site residues on WW1 for these two substrates. However, initial perturbations were detected for residues 322 Asn, 323 Thr and 327 Thr. Changes to residues in the predicted loop II region suggests that this may be the result of conformational changes in the presence of the peptide rather than direct interaction based on the previous identification of key residues in WW domain interactions. However, movement of 327 Thr located in the 3rd β -strand may be the result of binding by starting to stabilise the domain via interaction with this strand (Panwalker *et al.*, 2016; Panwalker *et al.*, 2017), although this speculation should be taken lightly given the minimal data available. Further investigation into these binding sites is required,

and additional experimentation such as the use of ITC may prove interesting in determining whether this domain has a preferred substrate like discovered for WW4. Another aspect that could be looked at as well is the phosphorylation status of the R-Smads, as WW4 presented a higher affinity for phospho-Smad 7. Although initial data has shown WW1 as the least folded and most unstable of the four domains of WWP2, while binding to substrates as previously described (Soond and Chantry, 2011a; Soond *et al.*, 2013), the addition of further experiments is essential in understanding binding and in turn therapeutic development towards this domain. This is of particular importance given the significant association of WWP2-N expression in early stages of prostate cancer.

To determine whether the problems faced with WW1 could be overcome by expression with the nearby WW2, and to investigate whether these domains act cooperatively, potentially altering the binding sites based on the presence and absence of the other domain, titration analysis was carried out using the WW1-2 tandem domain. The overlap of the end of WW1 and start of WW2 differs from between WW2 and WW3, linked to auto-inhibition (Chen *et al.*, 2017) and between WW3 and WW4, thought to regulate Smad 3 dependent gene activity (Yim, 2017). The reason for this is unclear, however as the other linker regions appear have a role in the regulation of WWP2, perhaps the lack of a linker suggests the potential for cross-communication between these first two domains, or even a concerted mode of action within WWP2-FL. The secondary structure prediction analysis (shown in section 4.3) for the tandem construct suggested that WW1 may not fold into the typical 3 stranded β -sheet in the presence of WW2 but could create an extended coil region in place of the 3rd strand. However, this should be taken lightly as no evidence has yet been uncovered relating to the structural solution of this tandem domain. Similar to HSQC analysis of WW1 and WW2 individually, the number of peaks present for the tandem domain did not correspond to the residue number for the construct, suggesting that amide peaks were missing, and like discussed earlier could be due to several factors either within the experimental design or inherent properties of the protein. Steps towards understanding this region as a potential cooperative tandem domain were carried out by investigating the interaction of these domains to Smad 2, 3 and 7 peptides. All titrations showed chemical shift perturbations indicative of peptide binding and domain folding. Of particular interest in these titrations were the changes in the tryptophan side-chains. As previously mentioned, the first tryptophan residue of WW domains has been associated with domain stability by interacting with proline in the hydrophobic core and the second

is part of the XP groove involved in substrate interaction. Within all three titrations with GB1-WW1-2, a change in chemical environment was observed for all four tryptophan ϵ -amide groups, but not those residing in GB1. This suggests that in the presence of Smad 2, 3 and 7 peptides, WW1-2 underwent conformational changes related to both stability and binding. There was also a slight difference between the tryptophan side chain positions with Smad 7 peptide compared to Smad 2 and 3 which were almost identical. This implies a different orientation to the side chains in the presence of R-smads and I-smad 7 and may be important in understanding how the binding differs between these Smad substrate types. As there is only a two-residue difference between the Smad 2 and 3 peptides, it was hypothesised that the binding site would be identical, and any differences between these interactions would be subtle. Without the assignments, it is difficult to conclude which amino acids in particular are being affected within the construct, however the potential appearance of arginine side chains suggests participation in hydrogen bonding with the peptide or possibly between domains. The majority of the peaks between the two titrations migrate in similar directions suggesting that it may be the cellular context that determines whether Smad 2 or Smad 3 is targeted. However, this is only a hypothesis based upon comparison of amide peak positions between the two titrations, and further assignments and analysis should uncover the regions and specific residues involved in binding. Furthermore, titration analysis of WW2 only with Smad 2 and 3 peptides should also be carried out to determine conclusively whether this domain can interact with these substrates as well. Understanding the differences between these two R-smads is also important in revealing the mechanism of selection between these substrates, although further work looking into the timing, rate and cellular context in the regulation of these interactions is required.

Interestingly, while research so far has suggested that WW1 and WW1-2 are responsible for R-smad binding and ubiquitination (Soond and Chantry, 2011a; Soond *et al.*, 2013), work by Tiffany Yim (2017) suggested that WW3 may in fact be the predominant interaction site for Smad 2 while WW1 is primarily responsible for Smad 3 binding. Overexpression of WW1-2 led to increased expression levels of Smad 3 whereas higher expression levels of Smad 2 were detected when WW3-4 was overexpressed. It would be intriguing to determine if there is such a binding preference between these domains, as shown for WW4 (Wahl, 2016), however other techniques may need to be employed for this such as ITC or fluorescence polarisation. This knowledge could then potentially help direct design towards one particular substrate,

as it would provide data on the structure-activity relationship (SAR). Often without this experimental data, software to predict such affinities are employed for rational peptide design which rely on prediction from the construct sequence or use data relating to the bound structure (Farhadi and Hashermain, 2018). Again, the latter requires the structural solution hence work such as the WW4 solution in chapter 3, and further work into solving the other domains, supports using this method for progressing peptide design as well as furthering biological understanding via protein structure. Collectively, substrate selection by these WW domains appears to be much more complex than originally thought, and the binding properties and biological function of these WW domains requires further analysis. The idea that these domains may not be easy to mimic without fully understanding their roles in both the full-length protein or as shorter isoforms and constructs further supports the need for more research.

Another perspective for inhibitor design other than target-based design discussed above is to inspect the other interaction partner – the Smad protein substrate – referred to as ligand-based peptide design approach. Using the information derived from such titrations, and structural data where possible, has shown only one side of the interaction. The smad proteins have a very clear PPxY motif required for interactions with WW domains, however most of the work in this thesis has looked at the protein-protein interactions from the WW domain perspective. While this is important in understanding the interaction, additional investigations using both unlabelled and isotopically labelled peptide would yield data regarding the specific residues within the substrate that contact the WW domain interaction partner, as well as reducing the number of potential residues needed to create a peptidomimetic. Unfortunately, efforts towards this were hindered by problems purifying Smad 2 and 3 peptides when recombinantly expressed thereby preventing such experimentation. However, Smad 7 purification (Wahl, 2016) and Oct4 peptide purification (Danielle Bourcier, 2018, personal communication) has shown that this should be possible but requires further design and optimisation. If such problems could be overcome, then smad-based peptides could be bacterially expressed in isotopically labelled minimal media and purified for use in future NMR experiments such that both unbound and bound conformations of the peptide and/or the interaction partner can be determined. One downside to this method is the problem obtaining labelled peptides that are phosphorylated to reflect such states during pathway activation. This is due to bacteria lacking the post-translational machinery required to phosphorylate these proteins. Within mammalian cells, phosphorylated Smad 2 and 3 are present by activation and

auto-phosphorylation of the TGF- β I receptor by interaction with TGF- β II receptor plus ligand binding leading to the activation of the kinase domain (Huse *et al.*, 2001; Xu *et al.*, 2012). In addition, for Smad 7 the kinase is currently unknown. Thus, replicating this process in bacterial cells is not currently possible. The alternatives for generating labelled and modified peptides are costly though as either the peptides require chemical synthesis with the addition of the phosphate group carried out after production, which in itself has complications in the accuracy of attachment to the correct atoms. Alternatively, the phospho-peptides can be made from individually labelled amino acids with the phosphate group already attached to the relevant residue(s). Support for this ligand-based structural approach in contrast to the target-based approach is also seen by the design of peptides using as few as 5 amino acids to cover an area of $\sim 1600 \text{ \AA}$ that are able to selectively inhibit protein-protein interactions (Sillerud and Larson, 2005). Once the difficulties in modified peptide generation are overcome, experiments similar to those carried out in the first two chapters of this thesis can be used to look at specific residues of the Smad-based peptides and their conformations when interacting with the different WW domains.

NMR is a powerful tool for facilitating structure-based rational design of peptide inhibitors as discussed above. However, NMR can also be used in the screening of small molecule inhibitors using a SAR-based directive (Skinner and Laurence, 2008). Similar to the titration experiments carried out in chapter 4, small ligands can be added to a sample in increasing concentrations to determine binding and kinetics. This would require a large abundance of isotopically labelled protein to enable the screening of many compounds with regards to WWP2, which as yet has proved difficult to obtain for these WW domains. As such, another direction was employed for small molecule screening.

6.3. Small-molecule based inhibitor discovery

Biochemical assay analysis of the E3 ligase enzyme activity was used to measure inhibition for the discovery of novel compounds against WWP2 activity. Assay development showed that the truncated isoform WWP2-C could not be used in the plate-assay set-up used here due to no detectable activity, unlike the similar screening assay by Rossi *et al.* (2014) using truncated Itch. One possible explanation for this is that WWP2-C may require the presence of substrates or an adaptor protein to activate its catalytic activity. Research by Mund and Pelham (2009) showed that WWP2 and other Nedd4 E3 ligases had increased auto-ubiquitination levels in the presence of the

adaptor protein NDFIP2. The analysis of Itch WW domain presence showed that WW2 had the most significant effect on NDFIP2-dependent auto-ubiquitination, suggesting that WW domain interaction with the PY motif is involved in the regulation of auto-ubiquitination. This is further supported by previous work with NDFIP1 where the presence of this adaptor protein relieved the WW domain/linker-mediated inhibition of Itch, WWP2 and Nedd4 activity (Riling *et al.*, 2015). This supports the activation of Nedd4 E3 ligases upon interaction with these adaptors and provides a potential basis for the lack of WWP2-C activity detected in this assay. However, this does not eliminate the possibility of other mechanisms used within the cellular environment to activate WWP2 isoforms. Recent work has suggested that the activation of members of the Nedd4 family may also be regulated by membrane interaction and polymerisation of substrates (Mund and Pelham, 2018). The evidence showed that upon mutation of key residues required for polymerisation in Dishevelled (Dvl), a substrate of WWP2, ubiquitination activity was abolished however the addition of a membrane interacting sequence to Dvl rescued ubiquitination. This supports these two conditions playing a role in WWP2 regulation. Yet, this research only investigated the full-length WWP2 with Dvl substrate therefore the application of this as a hypothesis for other isoforms and substrates of WWP2 can only be speculated and would require investigation.

The work by Chen and colleagues (2017) as mentioned previously which identified the WW2-3 linker as being a regulator of WWP2 activity contradicts the hypothesis that WWP2-C activation requires the presence of other proteins. Constructs used by Chen and colleagues included HECT only and WW3-WW4-HECT in which both presented with detectable auto-ubiquitination activity in western analysis, although the HECT only construct presented the highest degree of ubiquitination, potentially indicating constitutive activation. These initial experiments were carried out in solution in the absence of adaptor proteins or substrates rather than using GST-tag immobilisation of the E3 ligase. This comparison suggests that WWP2-C should not require the presence of an activator for auto-ubiquitination activity, and that the lack of activity in this assay may in fact be a result of the experimental set-up in which WWP2-C was immobilised to the plate. This therefore supports the hypothesis that WWP2-C may interact and dimerise with itself and/or other WWP2 isoforms to activate auto-ubiquitination. Interaction between isoforms has already been identified by Soond and Chantry (2011a) where WWP2-N showed interaction and activation of WWP2-FL. However, interaction combinations including WWP2-N: WWP2-N, WWP2-C: WWP2-C or WWP2-N: WWP2-C were not analysed in this work, thus it is currently unknown

whether other combinations of isoform dimerization exist. On the other hand, these investigations do not rule out the possibility that substrate presence may lead to increased activation of truncated isoforms, which was not tested in this research. Understanding more about the regulation of WWP2-C expression and activation in the cellular context to facilitate the use of these isoforms has however been difficult. Investigations into WWP2-C expression levels in Colo-357 cells by Soond and Chantry (2011a), and Tiffany Yim (Yim, 2017) in melanoma cell lines showed that WWP2-C expression levels were low or not detected when stimulated with TGF- β . However, Soond and Chantry showed that WWP2-C could be detected in ATDC5, a chondrogenic cell line derived from testis carcinoma which differentiates upon BMP signalling. Additionally, Tiffany Yim (Yim, 2017) also showed that expression of WWP2-C could be detected in few melanoma cell lines under BMP stimulation. This suggests that WWP2-C isoform activity on Smad 7, a common substrate in both pathways, may have a more significant effect in BMP signalling than TGF- β , however this requires investigation. Further experiments should include binding assays between WWP2 isoforms and the effect overexpression has on the activation and auto-ubiquitination of these isoforms to further understand the regulation of these proteins. This in turn may reveal information about the cellular contexts in which these proteins function to repress or stimulate oncogenesis, as well as facilitate the design of novel assays towards selective isoform inhibitor screening.

In contrast to the lack of significant activity by WWP2-C, the use of the full-length WWP2 was successful in the screening assay. Use of the full-length protein has the advantage of allowing the identification of small-molecules that inhibit activity via the binding to any region of WWP2, unlike currently known inhibitors of Nedd4 E3 ligases which have shown HECT domain specific inhibition (Kathman *et al.*, 2015; Mund *et al.*, 2014; Rossi *et al.*, 2014; Qirit *et al.*, 2017). Full-length WWP2 inhibitor screening may facilitate the development of small-molecule inhibitors against other isoforms of WWP2, by WW domain or linker binding. Although, given that the HECT domain is the catalytic site for ubiquitination by WWP2, it is more likely that inhibitors discovered in this screening assay interact with the HECT domain. The inclusion of the auto-inhibitory linker between WW2 and WW3 of WWP2 (Chen *et al.*, 2017) did not appear to have an effect on WWP2 auto-ubiquitination in these assays. This supports the idea that other aspects may be involved in the regulation of this region in inducing the auto-inhibitory conformation rather than the absence of substrate or adaptor proteins alone. While Nedd4 and WWP1 constructs used in this work for counter E3 ligase

screening showed auto-ubiquitination activity, interestingly when Smurf 2 was tested no auto-ubiquitination was detected (data not included). When taking into account auto-inhibitory conformations which have been reported for most Nedd4 family E3 ligases, this suggests that Smurf 2 adopts this inactive conformation as a result of substrate/adaptor absence. Although, the addition of Smad 7 peptide containing the PPxY motif failed to induce activation (data not included) suggesting that the auto-inhibition could not be relieved by the presence of the isolated Smad 7 PPxY motif region. This implies that adaptors such as NDFIPs may be the necessary factor for the activation of Smurf 2, or perhaps membrane association via the C2 domain of Smurf 2 is required. Although, the use of the full-length Smad 7 in breaking the C2-HECT interaction cannot be ruled out which has previously been shown to release this interaction (Wiesner *et al.*, 2007). This opens up further questions as to the cellular conditions that trigger auto-inhibition over auto-ubiquitination and turnover for different members of the Nedd4 family and such knowledge would be useful for further research involving selective inhibitor discovery as well as peptidomimetics which could mimic the inhibitory regions and block HECT activity.

Using WWP2-FL, plate-based screening of ~2100 compounds revealed 14 potential leads presenting IC₅₀ values < 11 μ M in dose-dependent analysis. This indicated good effectiveness at low concentrations, which is a desired property of drug leads as this is thought to result in a reduced number of undesired off-target effects. As these compounds have only been tested in *in vitro* auto-ubiquitination assays using purified proteins, it would be interesting to determine whether similar IC₅₀ values would be obtained in a cell-based ubiquitination assay. This would have the advantage of investigating activity levels of WWP2 in the absence and presence of the compounds in the cellular environment. But this may also reveal any problems associated with cellular solubility. This aspect has not been tested specifically in this work but is an important property for compounds that hope to be successful therapeutics. The variation in the structures of the selected hits indicated no obvious characteristic important for the inhibition observed in the screening data. Similarity was only observed between 228155 and 228150 showing similar indole-based rings with an NO₂ side group, and 288387 and 3064 with flavin-like structures. Collaboration with Serena Monaco and Sam Walpole (Watt *et al.*, 2018; unpublished observations) using ligand-based NMR revealed that 6 of the compounds discovered from the Diversity set, and 2 natural products, directly bound to the WWP2 HECT domain. Only one of the remaining 6 was deemed a non-binder. The others could not be determined possibly

due to solubility problems or from the binding of the ligand being too strong for detection. The non-binder was 228155, however this compound has previously been identified as a PAINS compound (Jasial, *et al.*, 2017), therefore this compound is most likely a false-positive result and not worth progressing further. The binding of the similar compound 228150 could not be determined, but due to its similarity to 228155 including this compound in future work beyond this thesis should be considered carefully.

As for the molecules that have been shown to interact with the HECT domain, further analysis has shown for a few examples that binding predominantly resides between the N- and C-lobes and has been hypothesised to block the flexible motions required for HECT domain function. With the use of STD epitope mapping and the recently developed DEEP-STD NMR by Monaco and colleagues (2017), Serena Monaco identified the residues within the HECT domain that contact the ligand as well as the atoms of the ligand that contact the protein. Using computational docking of the ligand 288387 to the WWP2-HECT domain, which was chosen based-on a medium binding affinity ($IC_{50} \sim 2.3 \mu M$), the modelling showed the phenyl ring of the compound resided deep inside the pocket between the two HECT lobes and makes contacts with hydrophobic leucine and proline residues (Watt *et al.*, 2018). The end of the flavin-like ring group is more solvent-exposed with contacts to valine and arginine by the opening of the pocket. It would be interesting to determine whether the similar compound 3064 makes the same or similar pattern on contacts that reflect binding between the lobes of the HECT domain as this has yet to be analysed. Using this approach for the other hits may be difficult though therefore for compounds unsuitable for NMR analysis, another method is required for binding site identification such as co-crystallisation as the unbound HECT domain has already been structurally solved. Despite some of the compounds may be potential PAINS and therefore should be used with caution, this is the first set of data for WWP2 inhibitors that shows low micromolar IC_{50} values, selectivity to the E3 ligases of the Nedd4 family, and has enabled the 3D mapping of the binding to the WWP2 HECT domain using a novel NMR approach. Binding analysis with other potential hits from this research may uncover further support for these compounds as potential inhibitors. Additionally, perhaps using other Nedd4 family HECT domains could reveal similarities and differences in binding that could be manipulated in further compound modification to increase the selectivity towards WWP2. Briefly mentioned in chapter 5, the differences in intramolecular interactions and sequences of the HECT domains may explain the different mechanisms revealed so far with other HECT domain inhibitors, and the variance in potency. Experimentation

to support the binding site between the lobes by either mutation of these key residues, or of other important sites within the HECT domain, for example the E2 binding site or the non-catalytic ubiquitin 'exosite' (Kamadurai *et al.*, 2009; Kim *et al.*, 2011; Maspero *et al.*, 2011) would help conclude that this region between the lobes is the predominant interaction site. This binding site between the lobes of the HECT domain has been speculated for the anti-depressant clomipramine that has been identified as an inhibitor of ubiquitin thiolation for the Nedd4 E3 ligase Itch by Rossi *et al.* (2014). In contrast to this, the Pelham lab (Mund *et al.*, 2014) published another HECT E3 ligase inhibitor called heclin that acted by oxidising the catalytic cysteine residue in the C-lobe although a lower degree of binding was detected to the hinge and N-lobe. This suggests that the binding sites for these compounds may be common between Nedd4 E3 ligases therefore modifications are required to increase selectivity. Interestingly, the difference between compounds I, II and III (heclin) (Mund *et al.*, 2014) was a methyl/ethyl group attached to the furan ring that seemed to increase the potency, however the variation in effect between the ligases tested reduced upon modification. This highlights the difficulties faced when modifying side groups that may result in an improvement of certain qualities but also reduce others. Despite the analogues of clomipramine showing varying effects on Itch, the effect the modifications had on other Nedd4 E3 ligases was not published therefore whether any of the modifications were able to increase potency and selectivity is unknown. One way to potentially increase specificity while maintaining or enhancing potency may be look into linking compounds that interact with different but nearby sites in the proteins with one providing higher selectivity and the other relaying the inhibitory effect or by linking E3 ligase selective inhibitory peptides to a compound for both potent inhibition and increased targeting of a particular ligase. This latter option has yet to be designed and tested however, but when taking into account the potential for inhibitory peptides generated based on Smad proteins as previously discussed, this may open up another area of research and potential drug design. Furthermore, knowledge of the mechanisms of action will also aid this directed design approach by providing information about whether the inhibitor is physically blocking, chemically changing the environment and/or acting upon another functional aspect of the protein.

For these compounds to progress into potential leads for therapeutic applications, their activity on the ubiquitination pathway had to be tested. As discussed in chapter 1, three enzymes are involved in the canonical ubiquitination pathway. In order to analyse the specificity of the compounds discovered in the initial screening towards the final E3

enzyme, counter assays using Uba1 (E1) and UbcH7 (E2) were carried out. Interestingly, as mentioned in chapter 5, no significant effect was observed in the E2 assay, but all compounds presented an inhibitory effect on E1 activity in the plate assay format experiment. The counter assay results suggest that these compounds do not affect the overall function of E1 and E2 steps in the ubiquitination cascade via the E2 assay results, however the partial inhibition observed in the E1 assay cannot be ignored. Further quantitative investigations into whether this is the result of the experimental conditions or if the compounds are having an inhibitory effect on E1 activity is required. The second assay used to observe E1 ubiquitin charging in the absence and presence of compound provided a method of visualising whether E1~Ub complexes formed in the samples after incubation with the test compounds. The presence of this complex in all samples suggested that none of the compounds were capable of complete inhibition but does not rule out a reduction in E1 activity as quantification of this method of counter screening was not carried out. Therefore, the results of the plate assay may be true, but require further work to determine this. In support of the latter possibility however, work using the E1 inhibitor PYR-41 showed partial inhibition of HECT E3 ligase activity in vitro (Yang *et al.*, 2007), suggesting that the plate assay results may reflect a similar partial inhibition mechanism as PYR-41, which blocks the E1 cysteine residue, between E1 and E3 enzymes. Furthermore, this may also suggest that these compounds have the potential to be modified such that they could become either E1 or E3 inhibitors. Further investigation into the binding sites and mechanism of action for the inhibitors screened here is required to determine whether this may be possible. Although, the direction of this work towards inhibitors against the E3 ligase WWP2 was of focus rather than E1 inhibitors. This is due to the increased selectivity of E3 ligases in targeting particular pathways, unlike E1 in which only 1 of 2 currently discovered in humans acts within the ubiquitination pathway.

When considering substrate ubiquitination, WWP2 has been found to target tumour suppressor proteins such as Smad 2, 3 and 7 within both the TGF- β signalling pathway and PTEN, a negative regulator of PI3K/AKT signalling, as outlined in chapter 1. WWP2 acts to ligate ubiquitin to the target lysine residues within these proteins in order for them to be recognised either for degradation or other cellular consequences (Chen *et al.*, 2016; Soond and Chantry, 2011b). The small molecule screening utilised auto-ubiquitination as a readout, and while this is adequate for initial discovery, the compounds required testing to establish if they could also inhibit substrate ubiquitination by WWP2. Results showed that the 7 compounds presenting IC₅₀ values

of 3 μM or lower all inhibited the ubiquitination of PTEN *in vitro*. This highlights that these compounds are functionally active against the activity of WWP2 towards this substrate. It is hypothesised that by inhibiting WWP2-mediated ubiquitination of PTEN using these hit compounds, this could result in the rescue of the PTEN steady state levels and in turn reduce AKT signalling by lowering the levels of phosphorylated AKT. These consequences are supported by work by Maddika *et al.* (2011) where WWP2 knockdown by siRNA in DU145 cells resulted in increased PTEN levels, reduction in phospho-AKT and increased sensitivity to stress-induced cell death. In contrast, overexpression of WWP2 resulted in the opposite occurring and cells developed resistance to apoptosis. The stabilisation of PTEN levels has also been reported in WWP2 knockout mice tissues (Li *et al.*, 2018). While the transcription levels of PTEN were not affected by WWP2 knockout, the evidence supports the role of WWP2 inhibition on a protein level at reducing the cancer profile facilitated by WWP2-mediated PTEN degradation. Interestingly, WWP2 knockout/knockdown did not affect the localisation of PTEN in both reports further supporting the role of WWP2 as a regulator of PTEN turnover. As PTEN insufficiency is a mediator of prostate cancer development, and WWP2 is commonly overexpressed in prostate cancer, inhibition of WWP2 using compounds identified here as a basis for further inhibitor optimisation, makes WWP2 a very promising druggable target. However, it is important to note that these potential inhibitors should also be tested for ubiquitination inhibition of other substrates to further validate these compounds and rule out non-specific activity or false-positives.

The overall aim of this work was to facilitate the development of novel inhibitors against WWP2 for potential use as an advanced or potentially earlier stage prostate cancer therapeutic. Current chemical treatment options, briefly mentioned in chapter 1, such as abiraterone, docetaxel and/or enzalutamide chemotherapy and hormone treatments, are limited to advanced cancer stages (Schweizer and Antonarakis, 2012), in which tumours either return, are no longer localised or are non-responsive to earlier therapies. This supports the requirement for further research and development of novel therapeutics. Initial experiments into the effect of the compounds with IC50 values below 3 μM was carried out in metastatic prostate cancer cell lines to establish whether these initial compounds had any effect on growth/migration. Preliminary findings showed that the presence of natural product 6 (401005) resulted in reduced growth/migration in both conditions in PC-3 cells but not DU145 cells. While compound 7 (228155) had an effect in this study, the PAINS identity of this compound suggests this is most likely non-specific, but it is still interesting that there is an effect on the PCa

cells by this compound and perhaps modifications may be able to overcome the PAIN compound aspects. While both cells lines are not responsive to androgens, PC-3 cells are highly metastatic compared to the moderately metastatic DU145 cells (ATCC CRL-1435TM; ATCC HTB-81TM). This suggests that natural product 6 may be showing more of an effect in more advanced prostate cancer. Other compounds seemed to present a TGF- β dependent effect in PC-3 cells. As to the cellular context of this result, no conclusions can be accurately drawn as this requires further experimentation. Furthermore, the ability for the compounds to penetrate the cellular membrane has not been tested, therefore the presence and concentration of the compounds within these cells is unknown. Further investigations into these compounds in cell-based experiments are necessary to further advance these molecules towards potential lead therapeutics.

6.4. Future work

- Obtain bound WW4 structures using isotopically labelled WW4 and labelled peptides to provide information relevant to understanding the interaction and enable interruption of the contacts. Docking studies could also be used to model peptide binding using the unbound refined structure.
- Design WW domain-based peptides and test alongside natural WW domains in auto-ubiquitination assays as well as non-cancerous and cancer cell-based assays, including the use of luciferase reporter assays to determine effect on TGF-B mediated transcription and signalling. Further experiments to investigate the effect of the compounds on PI3K/AKT signalling should also be considered
- Assign tandem domain HSQC peaks to determine residues important to interactions; collect Smad 2/3 titrations with WW2 to determine any differences between tandem WW1-2 and WW2 interactions and potentially reveal whether bound structures may be possible.
- Continue analysis, modification and optimisation of current compound hits and begin testing ADME; determine cellular effects of compounds in non-cancerous and cancerous conditions.
- Investigate the development of proteolysis-targeting chimera molecules (PROTACs) in which a WWP2 specific binding ligand is connected via a linker to another ligand either against another E3 ligase which is overexpressed in order to reduce the protein levels, or another oncogenic protein that is currently difficult to target using small molecule inhibition.

References

- Al-Khour, A. M. *et al.* (2005) 'Cooperative Phosphorylation of the Tumor Suppressor Phosphatase and Tensin Homologue (PTEN) by Casein Kinases and Glycogen Synthase Kinase 3 β ', *Journal of Biological Chemistry*, 280(42), pp. 35195–35202. doi: 10.1074/jbc.M503045200.
- Alarcón, C. *et al.* (2009) 'Nuclear CDKs Drive Smad Transcriptional Activation and Turnover in BMP and TGF- β Pathways', *Cell*, 139(4), pp. 757–769. doi: <https://doi.org/10.1016/j.cell.2009.09.035>.
- Aragón, E. *et al.* (2012) 'Structural basis for the versatile interactions of Smad7 with regulator WW domains in TGF- β pathways', *Structure*, 20(10), pp. 1726–1736. doi: 10.1016/j.str.2012.07.014.
- Aragón, E. *et al.* (2011) 'A smad action turnover switch operated by WW domain readers of a phosphoserine code', *Genes and Development*, 25(12), pp. 1275–1288. doi: 10.1101/gad.2060811.
- Aronchik, I. *et al.* (2014) 'The Antiproliferative Response of Indole-3-Carbinol in Human Melanoma Cells Is Triggered by an Interaction with NEDD4-1 and Disruption of Wild-Type PTEN Degradation', *Molecular Cancer Research*, 12(11), pp. 1621–1634. doi: 10.1158/1541-7786.MCR-14-0018.
- Baecker, V. (2011) MRI Wound healing tool, http://dev.mri.cnrs.fr/projects/imagej-macros/wiki/Wound_Healing_Tool, 2018.
- Baell, J. B. (2016) 'Feeling Nature's PAINS: Natural Products, Natural Product Drugs, and Pan Assay Interference Compounds (PAINS)', *Journal of Natural Products*. American Chemical Society, 79(3), pp. 616–628. doi: 10.1021/acs.jnatprod.5b00947.
- Baell, J. B. and Holloway, G. A. (2010) 'New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays', *Journal of Medicinal Chemistry*. American Chemical Society, 53(7), pp. 2719–2740. doi: 10.1021/jm901137j.
- Baryshnikova, O. K., Williams, T. C. and Sykes, B. D. (2008) 'Internal pH indicators for biomolecular NMR', *Journal of Biomolecular NMR*, 41(1), pp. 5–7. doi: 10.1007/s10858-008-9234-6.
- Bedford, M. T. *et al.* (2000) 'A Novel Pro-Arg Motif Recognized by WW Domains', *Journal of Biological Chemistry*, 275(14), pp. 10359–10369. doi: 10.1074/jbc.275.14.10359.
- Berman, H., Henrick, K. and Nakamura, H. (2003) 'Announcing the worldwide Protein Data Bank', *Nature Structural Biology*, p. 980. doi: 10.1038/nsb1203-980.
- Berman, H. *et al.* (2007) 'The worldwide Protein Data Bank (wwPDB): Ensuring a single, uniform archive of PDB data', *Nucleic Acids Research*, 35(SUPPL. 1). doi: 10.1093/nar/gkl971.
- Bermúdez Brito, M., Goulielmaki, E. and Papakonstanti, E. A. (2015) 'Focus on PTEN Regulation', *Frontiers in Oncology*. Frontiers Media S.A., 5, p. 166. doi: 10.3389/fonc.2015.00166.
- Bernassola, F. *et al.* (2008) 'The HECT Family of E3 Ubiquitin Ligases: Multiple Players in Cancer Development', *Cancer Cell*, pp. 10–21. doi: 10.1016/j.ccr.2008.06.001.

- Berndsen, C. E. and Wolberger, C. (2014) 'New insights into ubiquitin E3 ligase mechanism', *Nature Structural and Molecular Biology*, pp. 301–307. doi: 10.1038/nsmb.2780.
- Bhattacharya, A., Tejero, R. and Montelione, G. T. (2007) 'Evaluating protein structures determined by structural genomics consortia', *Proteins: Structure, Function and Genetics*, 66(4), pp. 778–795. doi: 10.1002/prot.21165.
- Birtle, A. J. *et al.* (2003) 'Clinical features of patients who present with metastatic prostate carcinoma and serum prostate-specific antigen (PSA) levels < 10 ng/mL', *Cancer*. Wiley-Blackwell, 98(11), pp. 2362–2367. doi: 10.1002/cncr.11821.
- Bitting, R. L. and Armstrong, A. J. (2013) 'Targeting the PI3K/Akt/mTOR pathway in castration-resistant prostate cancer', *Endocrine-Related Cancer*, 20(3), pp. R83–R99. doi: 10.1530/ERC-12-0394.
- Boase, N. A. and Kumar, S. (2015) 'NEDD4: The founding member of a family of ubiquitin-protein ligases', *Gene*, pp. 113–122. doi: 10.1016/j.gene.2014.12.020.
- Bobby, R. *et al.* (2013) 'Structure and dynamics of human Nedd4-1 WW3 in complex with the αENaC PY motif', *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1834(8), pp. 1632–1641. doi: 10.1016/j.bbapap.2013.04.031.
- Carnero, A. and Paramio, J. M. (2014) 'The PTEN/PI3K/AKT Pathway in vivo, Cancer Mouse Models', *Frontiers in Oncology*, p. 252. Available at: <https://www.frontiersin.org/article/10.3389/fonc.2014.00252>.
- Carver, B. S. *et al.* (2011) 'Reciprocal Feedback Regulation of PI3K and Androgen Receptor Signaling in PTEN-Deficient Prostate Cancer', *Cancer Cell*, 19(5), pp. 575–586. doi: 10.1016/j.ccr.2011.04.008.
- Chai, J. *et al.* (2003) 'Features of a Smad3 MH1-DNA complex: Roles of water and zinc in DNA binding', *Journal of Biological Chemistry*, 278(22), pp. 20327–20331. doi: 10.1074/jbc.C300134200.
- Chantry, A. (2011) 'WWP2 ubiquitin ligase and its isoforms: New biological insight and promising disease targets', *Cell Cycle*, pp. 2437–2439. doi: 10.4161/cc.10.15.16874.
- Charan, R. A. and LaVoie, M. J. (2015) 'Pathologic and therapeutic implications for the cell biology of parkin', *Molecular and Cellular Neuroscience*. doi: 10.1016/j.mcn.2015.02.008.
- Chen, C. *et al.* (2006) 'Ubiquitin E3 ligase WWP1 as an oncogenic factor in human prostate cancer', *Oncogene*. Nature Publishing Group, 26, p. 2386. Available at: <http://dx.doi.org/10.1038/sj.onc.1210021>.
- Chen, C. and Matesic, L. E. (2007) 'The Nedd4-like family of E3 ubiquitin ligases and cancer', *Cancer and Metastasis Reviews*, pp. 587–604. doi: 10.1007/s10555-007-9091-x.
- Chen, Z. *et al.* (2016) 'Molecular Features of PTEN Regulation by C-Terminal Phosphorylation', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M116.728980.
- Chen, Z. *et al.* (2017) 'A Tunable Brake for HECT Ubiquitin Ligases', *Molecular Cell*, 66(3), p. 345–357.e6. doi: 10.1016/j.molcel.2017.03.020.
- Chen, Z. *et al.* (2016) 'Enzymatic Analysis of PTEN Ubiquitylation by WWP2 and NEDD4-1 E3 Ligases', *Biochemistry*, 55(26), pp. 3658–3666. doi: 10.1021/acs.biochem.6b00448.

- Cheung, M. S. *et al.* (2010) 'DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure', *Journal of Magnetic Resonance*, 202(2), pp. 223–233. doi: 10.1016/j.jmr.2009.11.008.
- Chong, P. A. *et al.* (2010) 'Coupling of tandem Smad ubiquitination regulatory factor (Smurf) WW domains modulates target specificity', *Proceedings of the National Academy of Sciences*, 107(43), pp. 18404–18409. doi: 10.1073/pnas.1003023107.
- Chong, P. A. *et al.* (2006) 'An expanded WW domain recognition motif revealed by the interaction between Smad7 and the E3 ubiquitin ligase Smurf2', *Journal of Biological Chemistry*, 281(25), pp. 17069–17075. doi: 10.1074/jbc.M601493200.
- Ciccarese, C. *et al.* (2017) 'Prostate cancer heterogeneity: Discovering novel molecular targets for therapy', *Cancer Treatment Reviews*, pp. 68–73. doi: 10.1016/j.ctrv.2017.02.001.
- Ciechanover, A., Elias, S. and Heller, H. (1980) 'Characterization of the heatstable polypeptide of the ATP-dependent proteolytic system from reticulocytes', *Journal of Biological Chemistry*, 255(16), pp. 7525–7528.
- Ciehanover, A., Hod, Y. and Hershko, A. (1978) 'A heat-stable polypeptide component of an ATP-dependent proteolytic system from reticulocytes', *Biochemical and Biophysical Research Communications*, 81(4), pp. 1100–1105. doi: 10.1016/0006-291X(78)91249-4.
- Cole, C., Barber, J. D. and Barton, G. J. (2008) 'The Jpred 3 secondary structure prediction server', *Nucleic Acids Research*. Oxford University Press, 36(Web Server issue), pp. W197–W201. doi: 10.1093/nar/gkn238.
- Cong, W.-T. *et al.* (2010) 'A visible dye-based staining method for DNA in polyacrylamide gels by ethyl violet', *Analytical Biochemistry*, 402(1), pp. 99–101. doi: <https://doi.org/10.1016/j.ab.2010.03.017>.
- Cook, J. C. and Chock, P. B. (1992) 'Isoforms of mammalian ubiquitin-activating enzyme', *Journal of Biological Chemistry*, 267(34), pp. 24315–24321.
- Cooper, E. M. *et al.* (2004) 'Biochemical analysis of Angelman syndrome-associated mutations in the E3 ubiquitin ligase E6-associated protein', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M401302200.
- Cristofano, A. D. *et al.* (1998) 'Pten is essential for embryonic development and tumour suppression', *Nature Genetics*. Nature America Inc., 19, p. 348. Available at: <http://dx.doi.org/10.1038/1235>.
- Croxford, K. P. *et al.* (2017) 'The potential role of transforming growth factor beta family ligand interactions in prostate cancer', *AIMS Molecular Science*, 4(1), pp. 41–61. doi: 10.3934/molsci.2017.1.41.
- Cunningham, D. and You, Z. (2015) 'In vitro and in vivo model systems used in prostate cancer research', *Journal of biological methods*, 2(1), p. e17. doi: 10.14440/jbm.2015.63.
- David, C. J. and Massague, J. (2018) 'Contextual determinants of TGF beta action in development, immunity and cancer', *Nature Reviews Molecular Cell Biology*, 19(7), pp. 419–435. doi: 10.1038/s41580-018-0007-0.
- Delaglio, F. *et al.* (1995) 'NMRPipe: A multidimensional spectral processing system based on UNIX pipes', *Journal of Biomolecular NMR*, 6(3), pp. 277–293. doi: 10.1007/BF00197809.

- Demo, S. D. *et al.* (2007) 'Antitumor Activity of PR-171, a Novel Irreversible Inhibitor of the Proteasome', *Cancer Research*, 67(13), p. 6383 LP-6391. Available at: <http://cancerres.aacrjournals.org/content/67/13/6383.abstract>.
- Deng, L. *et al.* (2000) 'Activation of the I κ B kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain', *Cell*, 103(2), pp. 351–361. doi: 10.1016/S0092-8674(00)00126-4.
- Dennler, S. *et al.* (1998) 'Direct binding of Smad3 and Smad4 to critical TGF beta-inducible elements in the promoter of human plasminogen activator inhibitor-type 1 gene.', *The EMBO journal*, 17(11), pp. 3091–100. doi: 10.1093/emboj/17.11.3091.
- Dodson, E. J. *et al.* (2015) 'Versatile communication strategies among tandem WW domain repeats', *Experimental Biology and Medicine*, 240(3), pp. 351–360. doi: 10.1177/1535370214566558.
- Drozdetskiy, A. *et al.* (2015) 'JPred4: a protein secondary structure prediction server', *Nucleic Acids Research*. Oxford University Press, 43(Web Server issue), pp. W389–W394. doi: 10.1093/nar/gkv332.
- Dubrovskaya, A. *et al.* (2009) 'The role of PTEN/Akt/PI3K signaling in the maintenance and viability of prostate cancer stem-like cell populations', *Proceedings of the National Academy of Sciences*, 106(1), p. 268 LP-273. Available at: <http://www.pnas.org/content/106/1/268.abstract>.
- Ebisawa, T. *et al.* (2001) 'Smurf1 Interacts with Transforming Growth Factor- β Type I Receptor through Smad7 and Induces Receptor Degradation', *Journal of Biological Chemistry*, 276(16), pp. 12477–12480. doi: 10.1074/jbc.C100008200.
- Edelmann, M. J., Nicholson, B. and Kessler, B. M. (2011) 'Pharmacological targets in the ubiquitin system offer new ways of treating cancer, neurodegenerative disorders and infectious diseases', *Expert Reviews in Molecular Medicine*, 13, p. e35. doi: 10.1017/S1462399411002031.
- Espanel, X. and Sudol, M. (1999) 'A single point mutation in a Group I WW domain shifts its specificity to that of Group II WW domains', *Journal of Biological Chemistry*, 274(24), pp. 17284–17289. doi: 10.1074/jbc.274.24.17284.
- Fang, J. *et al.* (2007) 'PI3K/PTEN/AKT Signaling Regulates Prostate Tumor Angiogenesis', *Cellular signalling*, 19(12), pp. 2487–2497. doi: 10.1016/j.cellsig.2007.07.025.
- Farhadi, T. and Hashemian, S. M. (2018) 'Computer-aided design of amino acid-based therapeutics: a review', *Drug Design, Development and Therapy*. Dove Medical Press, 12, pp. 1239–1254. doi: 10.2147/DDDT.S159767.
- Fedoroff, O. Y. *et al.* (2004) 'The Structure and Dynamics of Tandem WW Domains in a Negative Regulator of Notch Signaling, Suppressor of Deltex', *Journal of Biological Chemistry*, 279(33), pp. 34991–35000. doi: 10.1074/jbc.M404987200.
- Fogh, R. H. *et al.* (2005) 'A framework for scientific data modeling and automated software development', *Bioinformatics*, 21(8), pp. 1678–1684. doi: 10.1093/bioinformatics/bti234.
- Fouladkou, F. *et al.* (2008) 'The ubiquitin ligase Nedd4-1 is dispensable for the regulation of PTEN stability and localization', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(25), pp. 8585–8590. doi: 10.1073/pnas.0803233105.

- French, M. E. *et al.* (2017) 'Mechanism of ubiquitin chain synthesis employed by a HECT domain ubiquitin ligase', *Journal of Biological Chemistry*, 292(25), pp. 10398–10413. doi: 10.1074/jbc.M117.789479.
- Fukumoto, C. *et al.* (2014) 'WWP2 is overexpressed in human oral cancer, determining tumor size and poor prognosis in patients: downregulation of WWP2 inhibits the AKT signaling and tumor growth in mice', *Oncoscience*, 1, p. 807. doi: 10.18632/oncoscience.101.
- Gao, S. *et al.* (2009) 'Ubiquitin Ligase Nedd4L Targets Activated Smad2/3 to Limit TGF- β Signaling', *Molecular Cell*, 36(3), pp. 457–468. doi: 10.1016/j.molcel.2009.09.043.
- Georgescu, M.-M. *et al.* (1999) 'The tumor-suppressor activity of PTEN is regulated by its carboxyl-terminal region', *Proceedings of the National Academy of Sciences of the United States of America*. The National Academy of Sciences, 96(18), pp. 10182–10187. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC17863/>.
- Gericke, A., Munson, M. and Ross, A. H. (2006) 'Regulation of the PTEN phosphatase', *Gene*, 374, pp. 1–9. doi: <https://doi.org/10.1016/j.gene.2006.02.024>.
- Gilberg, E. *et al.* (2016) 'Highly Promiscuous Small Molecules from Biological Screening Assays Include Many Pan-Assay Interference Compounds but Also Candidates for Polypharmacology', *Journal of Medicinal Chemistry*. American Chemical Society, 59(22), pp. 10285–10290. doi: 10.1021/acs.jmedchem.6b01314.
- Gomella, L. G., Petrylak, D. P. and Shayegan, B. (2014) 'Current management of advanced and castration resistant prostate cancer', *The Canadian journal of urology*, 21(2 Supp 1), pp. 1–6.
- Gong, W. *et al.* (2015) 'Structure of the HECT domain of human WWP2', *Acta Crystallographica Section F: Structural Biology Communications*, 71, pp. 1251–1257. doi: 10.1107/S2053230X1501554X.
- Goulet, C. C. *et al.* (1998) 'Inhibition of the epithelial Na⁺ channel by interaction of Nedd4 with a PY motif deleted in Liddle's syndrome', *Journal of Biological Chemistry*, 273(45), pp. 30012–30017. doi: 10.1074/jbc.273.45.30012.
- Graphpad Prism version 7 for Mac OS X (2017) USA, <http://www.graphpad.com>
- Groettrup, M. *et al.* (2008) 'Activating the ubiquitin family: UBA6 challenges the field', *Trends in Biochemical Sciences*, pp. 230–237. doi: 10.1016/j.tibs.2008.01.005.
- Gu, S. *et al.* (2018) 'PROTACs: An Emerging Targeting Technique for Protein Degradation in Drug Discovery', *BioEssays*. Wiley-Blackwell, 40(4), p. 1700247. doi: 10.1002/bies.201700247.
- Guldberg, P. *et al.* (1997) 'Disruption of the *MMAC1/PTEN* Gene by Deletion or Mutation Is a Frequent Event in Malignant Melanoma', *Cancer Research*, 57(17), p. 3660 LP-3663. Available at: <http://cancerres.aacrjournals.org/content/57/17/3660.abstract>.
- Haas, A. L. *et al.* (1982) 'Ubiquitin-activating enzyme. Mechanism and role in protein-ubiquitin conjugation.', *Journal of Biological Chemistry*, 257(5), pp. 2543–2548.
- Hammerich, K. H., Ayala, G. E. and Wheeler, T. M. (2008) 'Anatomy of the prostate gland and surgical pathology of prostate cancer', in *Prostate Cancer*, pp. 1–14. doi: 10.1017/CBO9780511551994.003.

- Harvey, K. F. *et al.* (1999) 'All three WW domains of murine Nedd4 are involved in the regulation of epithelial sodium channels by intracellular Na⁺', *Journal of Biological Chemistry*, 274(18), pp. 12525–12530. doi: 10.1074/jbc.274.18.12525.
- Harvey, K. F. and Kumar, S. (1999) 'Nedd4-like proteins: An emerging family of ubiquitin-protein ligases implicated in diverse cellular functions', *Trends in Cell Biology*, pp. 166–169. doi: 10.1016/S0962-8924(99)01541-X.
- Hata, A. *et al.* (1997) 'Mutations increasing autoinhibition inactivate tumour suppressors Smad2 and Smad4', *Nature*. Macmillan Magazines Ltd., 388, p. 82. Available at: <http://dx.doi.org/10.1038/40424>.
- Hayashi, H. *et al.* (1997) 'The MAD-Related Protein Smad7 Associates with the TGF β Receptor and Functions as an Antagonist of TGF β Signaling', *Cell*, 89(7), pp. 1165–1173. doi: [https://doi.org/10.1016/S0092-8674\(00\)80303-7](https://doi.org/10.1016/S0092-8674(00)80303-7).
- He, X. *et al.* (2012) 'PTEN Lipid Phosphatase Activity and Proper Subcellular Localization Are Necessary and Sufficient for Down-Regulating AKT Phosphorylation in the Nucleus in Cowden Syndrome', *The Journal of Clinical Endocrinology & Metabolism*, 97(11), pp. E2179–E2187. Available at: <http://dx.doi.org/10.1210/jc.2012-1991>.
- Hershko, A., Ciechanover, A. and Rose, I. A. (1979) 'Resolution of the ATP-dependent proteolytic system from reticulocytes: a component that interacts with ATP.', *Proceedings of the National Academy of Sciences*, 76(7), pp. 3107–3110. doi: 10.1073/pnas.76.7.3107.
- Huang, L. *et al.* (1999) 'Structure of an E6AP-UbcH7 complex: Insights into ubiquitination by the E2-E3 enzyme cascade', *Science*, 286(5443), pp. 1321–1326. doi: 10.1126/science.286.5443.1321.
- Huang, Y. J., Powers, R. and Montelione, G. T. (2005) 'Protein NMR recall, precision, and F-measure scores (RPF scores): Structure quality assessment measures based on information retrieval statistics', *Journal of the American Chemical Society*, 127(6), pp. 1665–1674. doi: 10.1021/ja047109h.
- Huibregtse, J. M. *et al.* (1995) 'A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase [published erratum appears in Proc Natl Acad Sci U S A 1995 May 23;92(11):5249]', *Proc Natl Acad Sci U S A*, 92(7), pp. 2563–2567.
- Huibregtse, J. M., Scheffner, M. and Howley, P. M. (1993) 'Localization of the E6-AP regions that direct human papillomavirus E6 binding, association with p53, and ubiquitination of associated proteins.', *Molecular and cellular biology*, 13(8), pp. 4918–27. doi: 10.1128/MCB.13.8.4918.Updated.
- Huse, M. *et al.* (2001) 'The TGF-beta Receptor Activation Process', *Molecular Cell*. Elsevier, 8(3), pp. 671–682. doi: 10.1016/S1097-2765(01)00332-X.
- Imoto, M. *et al.* (1990) 'Inhibition of EGF-induced phospholipase C activation in A341 cells by erbstatin, a tyrosine kinase inhibitor', *Biochemical and Biophysical Research Communications*, 173(1), pp. 208–211. doi: [https://doi.org/10.1016/S0006-291X\(05\)81042-3](https://doi.org/10.1016/S0006-291X(05)81042-3).
- Indig, G. L. (2002) 'Use of crystal violet as photochemotherapeutic agent.', *U.S. Pat. Appl. Publ.* USA.
- Ingham, R. J. *et al.* (2005) 'WW domains provide a platform for the assembly of multiprotein networks', *Mol Cell Biol*, 25(16), pp. 7092–7106. doi: 10.1128/MCB.25.16.7092-7106.2005.

- Itoh, Y. and Suzuki, M. (2018) 'Design, synthesis, and biological evaluation of novel ubiquitin-activating enzyme inhibitors', *Bioorganic and Medicinal Chemistry Letters*. doi: 10.1016/j.bmcl.2018.03.004.
- Jackson, P. K. *et al.* (2000) 'The lore of the RINGS: substrate recognition and catalysis by ubiquitin ligases', *Trends in Cell Biology*. Elsevier Current Trends, 10(10), pp. 429–439. doi: 10.1016/S0962-8924(00)01834-1.
- Jasial, S., Hu, Y. and Bajorath, J. (2017) 'How Frequently Are Pan-Assay Interference Compounds Active? Large-Scale Analysis of Screening Data Reveals Diverse Activity Profiles, Low Global Hit Frequency, and Many Consistently Inactive Compounds', *Journal of Medicinal Chemistry*. American Chemical Society, 60(9), pp. 3879–3886. doi: 10.1021/acs.jmedchem.7b00154.
- Jiang, J. *et al.* (2015) 'Characterization of substrate binding of the WW domains in human WWP2 protein', *FEBS Letters*, 589(15), pp. 1935–1942. doi: 10.1016/j.febslet.2015.05.021.
- Jolliffe, C. N. *et al.* (2000) 'Identification of multiple proteins expressed in murine embryos as binding partners for the WW domains of the ubiquitin-protein ligase Nedd4.', *Biochemical Journal*, 351(Pt 3), pp. 557–565. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1221394/>.
- Kaighn, M. E. *et al.* (1979) 'Establishment and characterization of a human prostatic carcinoma cell line (PC-3)', *Investigative urology*, 17(1), pp. 16–23. Available at: <http://europepmc.org/abstract/MED/447482>.
- Kamadurai, H. B. *et al.* (2013) 'Mechanism of ubiquitin ligation and lysine prioritization by a HECT E3', *eLife*, 2013(2). doi: 10.7554/eLife.00828.001.
- Kamadurai, H. B. *et al.* (2009) 'Insights into Ubiquitin Transfer Cascades from a Structure of a UbcH5B~Ubiquitin-HECTNEDD4L Complex', *Molecular Cell*, 36(6), pp. 1095–1102. doi: 10.1016/j.molcel.2009.11.010.
- Kanelis, V., Rotin, D. and Forman-Kay, J. D. (2001) 'Solution structure of a Nedd4 WW domain-ENaC peptide complex', *Nature Structural Biology*, 8(5), pp. 407–412. doi: 10.1038/87562.
- Kang, H. Y. *et al.* (2001) 'From transforming growth factor-beta signaling to androgen action: identification of Smad3 as an androgen receptor coregulator in prostate cancer cells.', *Proceedings of the National Academy of Sciences of the United States of America*, 98(6), pp. 3018–3023. doi: 10.1073/pnas.061305498.
- Karbowsky, M. and Youle, R. J. (2011) 'Regulating mitochondrial outer membrane proteins by ubiquitination and proteasomal degradation', *Current Opinion in Cell Biology*. doi: 10.1016/j.ceb.2011.05.007.
- Kasanov, J. *et al.* (2001) 'Characterizing Class I WW domains defines key specificity determinants and generates mutant domains with novel specificities', *Chemistry and Biology*, 8(3), pp. 231–241. doi: 10.1016/S1074-5521(01)00005-9.
- Kathman, S. G. *et al.* (2015) 'A Small Molecule That Switches a Ubiquitin Ligase from a Processive to a Distributive Enzymatic Mechanism', *Journal of the American Chemical Society*, 137(39), pp. 12442–12445. doi: 10.1021/jacs.5b06839.
- Kavsak, P. *et al.* (2000) 'Smad7 Binds to Smurf2 to Form an E3 Ubiquitin Ligase that Targets the TGF β Receptor for Degradation', *Molecular Cell*, 6(6), pp. 1365–1375. doi: [https://doi.org/10.1016/S1097-2765\(00\)00134-9](https://doi.org/10.1016/S1097-2765(00)00134-9).

- Keeler, J. (2010) '*Understanding NMR Spectroscopy*', 2nd edition, John Wiley & Sons, Inc.
- Khan, M. I. *et al.* (2015) 'Role of Epithelial Mesenchymal Transition in Prostate Tumorigenesis', *Current pharmaceutical design*, 21(10), pp. 1240–1248. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4362522/>.
- Kim, H. C. *et al.* (2011) 'Structure and function of a HECT domain ubiquitin-binding site', *EMBO Reports*, 12(4), pp. 334–341. doi: 10.1038/embor.2011.23.
- Kirkin, V. and Dikic, I. (2007) 'Role of ubiquitin- and Ubl-binding proteins in cell signaling', *Current Opinion in Cell Biology*, pp. 199–205. doi: 10.1016/j.ceb.2007.02.002.
- Knopf, J. L. *et al.* (1986) 'Cloning and expression of multiple protein kinase C cDNAs', *Cell*, 46(4), pp. 491–502. doi: [https://doi.org/10.1016/0092-8674\(86\)90874-3](https://doi.org/10.1016/0092-8674(86)90874-3).
- Komander, D. and Rape, M. (2012) 'The Ubiquitin Code', *Annual Review of Biochemistry*, 81(1), pp. 203–229. doi: 10.1146/annurev-biochem-060310-170328.
- Komuro, A. *et al.* (2004) 'Negative regulation of transforming growth factor-beta (TGF-beta) signaling by WW domain-containing protein 1 (WWP1).', *Oncogene*, 23(41), pp. 6914–6923. doi: 10.1038/sj.onc.1207885.
- Kruse, J. P. and Gu, W. (2009) 'Modes of p53 Regulation', *Cell*, pp. 609–622. doi: 10.1016/j.cell.2009.04.050.
- Kuratomi, G. *et al.* (2005) 'NEDD4-2 (neural precursor cell expressed, developmentally down-regulated 4-2) negatively regulates TGF- β (transforming growth factor- β) signalling by inducing ubiquitin-mediated degradation of Smad2 and TGF- β type I receptor', *Biochemical Journal*. Portland Press Ltd., 386(Pt 3), pp. 461–470. doi: 10.1042/BJ20040738.
- Landré, V. *et al.* (2014) 'Screening for E3-ubiquitin ligase inhibitors: challenges and opportunities.', *Oncotarget*, 5(18), pp. 7988–8013. doi: 10.18632/oncotarget.2431.
- Laskowski, R. A. *et al.* (1993) 'PROCHECK: a program to check the stereochemical quality of protein structures', *Journal of Applied Crystallography*, 26(2), pp. 283–291. doi: 10.1107/S0021889892009944.
- Lau, J. L. and Dunn, M. K. (2018) 'Therapeutic peptides: Historical perspectives, current development trends, and future directions', *Bioorganic & Medicinal Chemistry*, 26(10), pp. 2700–2707. doi: <https://doi.org/10.1016/j.bmc.2017.06.052>.
- Lee, I. and Schindelin, H. (2008) 'Structural Insights into E1-Catalyzed Ubiquitin Activation and Transfer to Conjugating Enzymes', *Cell*, 134(2), pp. 268–278. doi: 10.1016/j.cell.2008.05.046.
- Lee, J. R. R. *et al.* (2009) 'The HECT domain of the ubiquitin ligase Rsp5 contributes to substrate recognition', *Journal of Biological Chemistry*, 284(46), pp. 32126–32137. doi: 10.1074/jbc.M109.048629.
- Lee, J.-O. *et al.* (1999) 'Crystal Structure of the PTEN Tumor Suppressor', *Cell*, 99(3), pp. 323–334. doi: 10.1016/S0092-8674(00)81663-3.
- Lee, J. T., Steelman, L. S. and McCubrey, J. A. (2004) 'Phosphatidylinositol 3'-kinase activation leads to multidrug resistance protein-1 expression and subsequent chemoresistance in advanced prostate cancer cells', *Cancer Research*, 64(22), pp. 8397–8404. doi: 10.1158/0008-5472.CAN-04-1612.

- Leite, K. R. M. *et al.* (2001) 'Abnormal expression of MDM2 in prostate carcinoma', *Modern Pathology*, 14(5), pp. 428–436. doi: 10.1038/modpathol.3880330.
- Lewis, M. R. and Goland, P. P. (1953) 'The tumor-inhibitory activity of diaryl- and triarylmethane dyes. I. The Ehrlich ascites mouse tumor.', *Cancer Research*, 13, pp. 130–136.
- Li, H. *et al.* (2018) 'WWP2 is a physiological ubiquitin ligase for phosphatase and tensin homolog (PTEN) in mice', *Journal of Biological Chemistry*, 293(23), pp. 8886–8889. doi: 10.1074/jbc.RA117.001060.
- Lim, K. L. *et al.* (2005) 'Parkin mediates nonclassical, proteasomal-independent ubiquitination of synphilin-1: implications for Lewy body formation', *Journal of Neuroscience*, 25(8), pp. 2002–9. doi: 10.1523/JNEUROSCI.4474-04.2005
- Lin, X., Liang, M. and Feng, X.-H. (2000) 'Smurf2 Is a Ubiquitin E3 Ligase Mediating Proteasome-dependent Degradation of Smad2 in Transforming Growth Factor- β Signaling', *Journal of Biological Chemistry*, 275(47), pp. 36818–36822. doi: 10.1074/jbc.C000580200.
- Linge, J. P. and Nilges, M. (1999) 'Influence of non-bonded parameters on the quality of NMR structures: A new force field for NMR structure calculation', *Journal of Biomolecular NMR*, 13(1), pp. 51–59. doi: 10.1023/A:1008365802830.
- Linge, J. P., O'Donoghue, S. I. and Nilges, M. (2001) 'Automated assignment of ambiguous nuclear overhauser effects with ARIA', *Methods in Enzymology*, 339, pp. 71–90. doi: 10.1016/S0076-6879(01)39310-2.
- Linge, J. P. *et al.* (2003) 'ARIA: Automated NOE assignment and NMR structure calculation', *Bioinformatics*, 19(2), pp. 315–316. doi: 10.1093/bioinformatics/19.2.315.
- Linge, J. P. *et al.* (2003) 'Refinement of protein structures in explicit solvent', *Proteins: Structure, Function and Genetics*, 50(3), pp. 496–506. doi: 10.1002/prot.10299.
- Livneh, I. *et al.* (2017) 'Monoubiquitination joins polyubiquitination as an esteemed proteasomal targeting signal', *BioEssays*. doi: 10.1002/bies.201700027.
- Lorenz, S. (2018) 'Structural mechanisms of HECT-type ubiquitin ligases', *Biological Chemistry*, pp. 127–145. doi: 10.1515/hsz-2017-0184.
- Lott, J. S. *et al.* (2002) 'A single WW domain is the predominant mediator of the interaction between the human ubiquitin-protein ligase Nedd4 and the human epithelial sodium channel', *Biochemical Journal*, 361, pp. 481–488. doi: 10.1042/0264-6021:3610481.
- Lovell, S. C. *et al.* (2003) 'Structure validation by C alpha geometry: phi,psi and C beta deviation', *Proteins-Structure Function and Genetics*, 50(August 2002), pp. 437–450. doi: 10.1002/prot.10286.
- Lüthy, R., Bowie, J. U. and Eisenberg, D. (1992) 'Assessment of protein models with three-dimensional profiles', *Nature*, 356(6364), pp. 83–85. doi: 10.1038/356083a0.
- Macias, M. J. *et al.* (1996) 'Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide', *Nature*, 382(6592), pp. 646–649. doi: 10.1038/382646a0.
- Macias, M. J., Martin-Malpartida, P. and Massagué, J. (2015) 'Structural determinants of Smad function in TGF- β signaling', *Trends in Biochemical Sciences*, pp. 296–308. doi: 10.1016/j.tibs.2015.03.012.

- Macias, M. J., Wiesner, S. and Sudol, M. (2002) 'WW and SH3 domains, two different scaffolds to recognize proline-rich ligands', *FEBS Letters*, pp. 30–37. doi: 10.1016/S0014-5793(01)03290-2.
- Maddika, S. *et al.* (2011) 'WWP2 is an E3 ubiquitin ligase for PTEN', *Nature Cell Biology*, 13(6), pp. 728–733. doi: 10.1038/ncb2240.
- Malakhov, M. P. *et al.* (2004) 'SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins', *Journal of Structural and Functional Genomics*, 5(1), pp. 75–86. doi: 10.1023/B:JSFG.0000029237.70316.52.
- Mari, S. *et al.* (2014) 'Structural and functional framework for the autoinhibition of nedd4-family ubiquitin ligases', *Structure*, 22(11), pp. 1639–1649. doi: 10.1016/j.str.2014.09.006.
- Markovits, J. *et al.* (1994) 'Inhibition of DNA topoisomerases I and II and induction of apoptosis by erbstatin and tyrphostin derivatives', *Biochemical Pharmacology*, 48(3), pp. 549–560. doi: [https://doi.org/10.1016/0006-2952\(94\)90285-2](https://doi.org/10.1016/0006-2952(94)90285-2).
- Maspero, E. *et al.* (2011) 'Structure of the HECT:ubiquitin complex and its role in ubiquitin chain elongation', *EMBO Reports*, 12(4), pp. 342–349. doi: 10.1038/embor.2011.21.
- Maspero, E. *et al.* (2013) 'Structure of a ubiquitin-loaded HECT ligase reveals the molecular basis for catalytic priming', *Nature Structural and Molecular Biology*, 20(6), pp. 696–701. doi: 10.1038/nsmb.2566.
- McDonald, F. J. *et al.* (2002) 'Ubiquitin-protein ligase WWP2 binds to and downregulates the epithelial Na⁺ channel', *American Journal of Physiology-Renal Physiology*, 283(3), pp. F431–F436. doi: 10.1152/ajprenal.00080.2002.
- Meierhofer, D. and Kaiser, P. (2011) 'Global analysis of ubiquitination', *Neuromethods*, 57, pp. 197–209. doi: 10.1007/978-1-61779-111-6_15.
- Metzger, M. B. *et al.* (2014) 'RING-type E3 ligases: Master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination', *Biochimica et Biophysica Acta - Molecular Cell Research*, pp. 47–60. doi: 10.1016/j.bbamcr.2013.05.026.
- Miller, S. J. *et al.* (2002) 'Direct identification of PTEN phosphorylation sites', *FEBS Letters*, 528(1), pp. 145–153. doi: [https://doi.org/10.1016/S0014-5793\(02\)03274-X](https://doi.org/10.1016/S0014-5793(02)03274-X).
- Monaco, S. *et al.* (2017) 'Differential Epitope Mapping by STD NMR Spectroscopy To Reveal the Nature of Protein–Ligand Contacts', *Angewandte Chemie - International Edition*, 56(48), pp. 15289–15293. doi: 10.1002/anie.201707682.
- Montanari, M. *et al.* (2017) 'Epithelial-mesenchymal transition in prostate cancer: an overview', *Oncotarget*. Impact Journals LLC, 8(21), pp. 35376–35389. doi: 10.18632/oncotarget.15686.
- Morén, A. *et al.* (2005) 'Degradation of the Tumor Suppressor Smad4 by WW and HECT Domain Ubiquitin Ligases', *Journal of Biological Chemistry*, 280(23), pp. 22115–22123. doi: 10.1074/jbc.M414027200.
- Mund, T. *et al.* (2014) 'Peptide and small molecule inhibitors of HECT-type ubiquitin ligases', *Proceedings of the National Academy of Sciences*, 111(47), pp. 16736–16741. doi: 10.1073/pnas.1412152111.
- Mund, T. and Pelham, H. R. B. (2009) 'Control of the activity of WW-HECT domain E3 ubiquitin ligases by NDFIP proteins', *EMBO Reports*. Nature Publishing Group, 10(5), pp. 501–507. doi: 10.1038/embor.2009.30.

- Mund, T. and Pelham, H. R. (2018) 'Substrate clustering potently regulates the activity of WW-HECT domain-containing ubiquitin ligases', *Journal of Biological Chemistry*, 293(14), pp. 5200–5209. doi: 10.1074/jbc.RA117.000934.
- Nakagawa, T. and Nakayama, K. (2015) 'Protein monoubiquitylation: Targets and diverse functions', *Genes to Cells*, pp. 543–562. doi: 10.1111/gtc.12250.
- Nakagawa, T. *et al.* (2008) 'Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation', *Genes and Development*, 22(1), pp. 37–49. doi: 10.1101/gad.1609708.
- Nan, B. *et al.* (2003) 'The PTEN tumor suppressor is a negative modulator of androgen receptor transcriptional activity', *Journal of Molecular Endocrinology*, 31(1), pp. 169–183. doi: 10.1677/jme.0.0310169.
- Nathan, J. A. *et al.* (2013) 'Why do cellular proteins linked to K63-polyubiquitin chains not associate with proteasomes?', *EMBO Journal*, 32(4), pp. 552–565. doi: 10.1038/emboj.2012.354.
- Nilges, M. (1997) 'Ambiguous distance data in the calculation of NMR structures', *Folding and Design*, 2(4). doi: 10.1016/S1359-0278(97)00064-3.
- Nilges, M. *et al.* (2008) 'Accurate NMR Structures Through Minimization of an Extended Hybrid Energy', *Structure*, 16(9), pp. 1305–1312. doi: 10.1016/j.str.2008.07.008.
- Nilges, M. *et al.* (1997) 'Automated NOESY interpretation with ambiguous distance restraints: The refined NMR solution structure of the pleckstrin homology domain from β -spectrin', *Journal of Molecular Biology*, 269(3), pp. 408–422. doi: 10.1006/jmbi.1997.1044.
- Nilges, M. and O'Donoghue, S. I. (1998) 'Ambiguous NOEs and automated NOE assignment', *Progress in Nuclear Magnetic Resonance Spectroscopy*, 32(2), pp. 107–139. doi: 10.1016/S0079-6565(97)00025-3.
- Odriezola, L. *et al.* (2007) 'Regulation of PTEN activity by its carboxyl-terminal autoinhibitory domain', *Journal of Biological Chemistry*, 282(32), pp. 23306–23315. doi: 10.1074/jbc.M611240200.
- Ogunjimi, A. A. *et al.* (2005) 'Regulation of Smurf2 ubiquitin ligase activity by anchoring the E2 to the HECT domain', *Molecular Cell*, 19(3), pp. 297–308. doi: 10.1016/j.molcel.2005.06.028.
- Oliver, P. M. *et al.* (2006) 'Ndfip1 Protein Promotes the Function of Itch Ubiquitin Ligase to Prevent T Cell Activation and T Helper 2 Cell-Mediated Inflammation', *Immunity*, 25(6), pp. 929–940. doi: 10.1016/j.immuni.2006.10.012.
- Orlowski, R. Z. *et al.* (2002) 'Phase I Trial of the Proteasome Inhibitor PS-341 in Patients With Refractory Hematologic Malignancies', *Journal of Clinical Oncology*. American Society of Clinical Oncology, 20(22), pp. 4420–4427. doi: 10.1200/JCO.2002.01.133.
- Packer, J. and Maitland, N. (2016) 'The molecular and cellular origin of human prostate cancer', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, pp. 1238–1260. doi: 10.1016/j.bbamcr.2016.02.016.
- Panwalker, V. *et al.* (2016) 'The Nedd4-1 WW Domain Recognizes the PY Motif Peptide through Coupled Folding and Binding Equilibria', *Biochemistry*, 55(4), pp. 659–674. doi: 10.1021/acs.biochem.5b01028.

- Panwalker, V. *et al.* (2017) 'Multiple WW domains of Nedd4-1 undergo conformational exchange that is quenched upon peptide binding', *FEBS Letters*. Wiley-Blackwell, 591(11), pp. 1573–1583. doi: 10.1002/1873-3468.12664.
- Pelzer, C. *et al.* (2007) 'UBE1L2, a novel E1 enzyme specific for ubiquitin', *Journal of Biological Chemistry*, 282(32), pp. 23010–23014. doi: 10.1074/jbc.C700111200.
- Pettersen, E. F. *et al.* (2004) 'UCSF Chimera', *Journal of Computational Chemistry*, pp. 1605–12. doi: 10.1002/jcc.20084.
- Pichler, A. *et al.* (2005) 'SUMO modification of the ubiquitin-conjugating enzyme E2-25K', *Nature Structural and Molecular Biology*, 12(3), pp. 264–9. Epub 2005 Feb 20.10.1038/nsmb903
- Pickart, C. M. (2001) 'Mechanisms Underlying Ubiquitination', *Annual Review of Biochemistry*, 70(1), pp. 503–533. doi: 10.1146/annurev.biochem.70.1.503.
- Pickart, C. M. and Eddins, M. J. (2004) 'Ubiquitin: Structures, functions, mechanisms', *Biochimica et Biophysica Acta - Molecular Cell Research*, pp. 55–72. doi: 10.1016/j.bbamcr.2004.09.019.
- Pirozzi, G. *et al.* (1997) 'Identification of novel human WW domain-containing proteins by cloning of ligand targets', *Journal of Biological Chemistry*, 272(23), pp. 14611–14616. doi: 10.1074/jbc.272.23.14611.
- Plant, P. J. *et al.* (1997) 'The C2 domain of the ubiquitin protein ligase Nedd4 mediates Ca²⁺- dependent plasma membrane localization', *Journal of Biological Chemistry*, 272(51), pp. 32329–32336. doi: 10.1074/jbc.272.51.32329.
- Platzer, G., Okon, M. and McIntosh, L. P. (2014) 'pH-dependent random coil ¹H, ¹³C, and ¹⁵N chemical shifts of the ionizable amino acids: a guide for protein pK a measurements', *Journal of Biomolecular NMR*, 60(2–3), pp. 109–129. doi: 10.1007/s10858-014-9862-y.
- Plechanovov, A. *et al.* (2012) 'Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis', *Nature*, 489(7414), pp. 115–120. doi: 10.1038/nature11376.
- Pruneda, J. N. *et al.* (2012) 'Structure of an E3:E2~Ub Complex Reveals an Allosteric Mechanism Shared among RING/U-box Ligases', *Molecular Cell*, 47(6), pp. 933–942. doi: 10.1016/j.molcel.2012.07.001.
- Qi, S. *et al.* (2014) 'Structural and Biochemical Basis for Ubiquitin Ligase Recruitment by Arrestin-related Domain-containing Protein-3 (ARRDC3)', *The Journal of Biological Chemistry*. 9650 Rockville Pike, Bethesda, MD 20814, U.S.A.: American Society for Biochemistry and Molecular Biology, 289(8), pp. 4743–4752. doi: 10.1074/jbc.M113.527473.
- Quirit, J. G. *et al.* (2017) 'Indole-3-carbinol (I3C) analogues are potent small molecule inhibitors of NEDD4-1 ubiquitin ligase activity that disrupt proliferation of human melanoma cells', *Biochemical Pharmacology*, 127, pp. 13–27. doi: 10.1016/j.bcp.2016.12.007.
- Ramaswamy, S. *et al.* (1999) 'Regulation of G1 progression by the PTEN tumor suppressor protein is linked to inhibition of the phosphatidylinositol 3-kinase/Akt pathway', *Proceedings of the National Academy of Sciences*, 96(5), p. 2110 LP-2115. Available at: <http://www.pnas.org/content/96/5/2110>.

- Richardson, P. G. *et al.* (2003) 'A Phase 2 Study of Bortezomib in Relapsed, Refractory Myeloma', *New England Journal of Medicine*. Massachusetts Medical Society, 348(26), pp. 2609–2617. doi: 10.1056/NEJMoa030288.
- Richardson, P. G. *et al.* (2005) 'Bortezomib or High-Dose Dexamethasone for Relapsed Multiple Myeloma', *New England Journal of Medicine*. Massachusetts Medical Society, 352(24), pp. 2487–2498. doi: 10.1056/NEJMoa043445.
- Rieping, W. *et al.* (2007) 'ARIA2: Automated NOE assignment and data integration in NMR structure calculation', *Bioinformatics*, 23(3), pp. 381–382. doi: 10.1093/bioinformatics/btl589.
- Rieser, E., Cordier, S. M. and Walczak, H. (2013) 'Linear ubiquitination: A newly discovered regulator of cell signalling', *Trends in Biochemical Sciences*, pp. 94–102. doi: 10.1016/j.tibs.2012.11.007.
- Riley, B. E. *et al.* (2013) 'Structure and function of Parkin E3 ubiquitin ligase reveals aspects of RING and HECT ligases', *Nature Communications*. doi: 10.1038/ncomms2982.
- Riling, C. *et al.* (2015) 'Itch WW domains inhibit its E3 ubiquitin ligase activity by blocking E2-E3 ligase trans-thiolation', *Journal of Biological Chemistry*, 290(39), pp. 23875–23887. doi: 10.1074/jbc.M115.649269.
- Ronchi, V. P. *et al.* (2014) 'The Active Form of E6-associated protein (E6AP)/UBE3A Ubiquitin Ligase Is an Oligomer', *The Journal of Biological Chemistry*. 9650 Rockville Pike, Bethesda, MD 20814, U.S.A.: American Society for Biochemistry and Molecular Biology, 289(2), pp. 1033–1048. doi: 10.1074/jbc.M113.517805.
- Rossi, M. *et al.* (2014) 'High throughput screening for inhibitors of the HECT ubiquitin E3 ligase ITCH identifies antidepressant drugs as regulators of autophagy', *Cell Death and Disease*, 5(5). doi: 10.1038/cddis.2014.113.
- Rossi, M. *et al.* (2005) 'The ubiquitin-protein ligase Itch regulates p73 stability', *EMBO Journal*, 24(4), pp. 836–848. doi: 10.1038/sj.emboj.7600444.
- Schäfer, A., Kuhn, M. and Schindelin, H. (2014) 'Structure of the ubiquitin-activating enzyme loaded with two ubiquitin molecules', *Acta Crystallographica Section D: Biological Crystallography*, 70(5), pp. 1311–1320. doi: 10.1107/S1399004714002910.
- Scheffner, M. and Kumar, S. (2014) 'Mammalian HECT ubiquitin-protein ligases: Biological and pathophysiological aspects', *Biochimica et Biophysica Acta - Molecular Cell Research*, 1843(1), pp. 61–74. doi: 10.1016/j.bbamcr.2013.03.024.
- Schindelin, J. *et al.* (2012) 'Fiji - an Open Source platform for biological image analysis', *Nature methods*, 9(7), p. 10.1038/nmeth.2019. doi: 10.1038/nmeth.2019.
- Schlesinger, D. H., Goldstein, G. and Niall, H. D. (1975) 'The Complete Amino Acid Sequence of Ubiquitin, an Adenylate Cyclase Stimulating Polypeptide Probably Universal in Living Cells', *Biochemistry*, 14(10), pp. 2214–2218. doi: 10.1021/bi00681a026.
- Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012) 'NIH Image to ImageJ: 25 years of Image Analysis', *Nature methods*, 9(7), pp. 671–675. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5554542/>.
- Schulman, B. A. and Wade Harper, J. (2009) 'Ubiquitin-like protein activation by E1 enzymes: The apex for downstream signalling pathways', *Nature Reviews Molecular Cell Biology*, pp. 319–331. doi: 10.1038/nrm2673.

- Schweizer, M. T. and Antonarakis, E. S. (2012) 'Abiraterone and other novel androgen-directed strategies for the treatment of prostate cancer: a new era of hormonal therapies is born', *Therapeutic Advances in Urology*. Sage UK: London, England: SAGE Publications, 4(4), pp. 167–178. doi: 10.1177/1756287212452196.
- Sillerud, L. O. and Larson, R. S. (2005) 'Design and structure of peptide and peptidomimetic antagonists of protein-protein interaction.', *Current protein & peptide science*, 6(2), pp. 151–169. doi: 10.2174/1389203053545462.
- Singh Parihar, J. and Yi Kim, I. (2016) 'Chapter 57 - Second-Line Hormonal for Castrate-Resistant Prostate Cancer', in Mydlo, J. H. and Godec, C. J. B. T.-P. C. (Second E. (eds). San Diego: Academic Press, pp. 533–540. doi: <https://doi.org/10.1016/B978-0-12-800077-9.00057-8>.
- Singh, S. *et al.* (2003) 'Overexpression of Vimentin', *Cancer Research*, 63(9), p. 2306 LP-2311. Available at: <http://cancerres.aacrjournals.org/content/63/9/2306.abstract>.
- Sippl, M. J. (1993) 'Recognition of Errors in 3-Dimensional Structures of Proteins', *Proteins-Structure Function and Genetics*, 17(4), pp. 355–362. doi: 10.1002/prot.340170404.
- Skinner, A. L. and Laurence, J. S. (2008) 'High-field Solution NMR Spectroscopy as a Tool for Assessing Protein Interactions with Small Molecule Ligands', *Journal of pharmaceutical sciences*, 97(11), pp. 4670–4695. doi: 10.1002/jps.21378.
- Skinner, S. P. *et al.* (2015) 'Structure calculation, refinement and validation using CcpNmr Analysis', *Acta Crystallographica Section D: Biological Crystallography*, 71, pp. 154–161. doi: 10.1107/S1399004714026662.
- Song, M. S., Salmena, L. and Pandolfi, P. P. (2012) 'The functions and regulation of the PTEN tumour suppressor', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 13, p. 283. Available at: <http://dx.doi.org/10.1038/nrm3330>.
- Soond, S. M. and Chantry, A. (2011a) 'Selective targeting of activating and inhibitory Smads by distinct WWP2 ubiquitin ligase isoforms differentially modulates TGF β signalling and EMT', *Oncogene*, 30(21), pp. 2451–2462. doi: 10.1038/onc.2010.617.
- Soond, S. M. and Chantry, A. (2011b) 'How ubiquitination regulates the TGF- β signalling pathway: New insights and new players: New isoforms of ubiquitin-activating enzymes in the E1-E3 families join the game', *BioEssays*, 33(10), pp. 749–758. doi: 10.1002/bies.201100057.
- Soond, S. M. *et al.* (2013) 'Novel WWP2 ubiquitin ligase isoforms as potential prognostic markers and molecular targets in cancer', *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1832(12), pp. 2127–2135. doi: 10.1016/j.bbadis.2013.08.001.
- Staub, O. *et al.* (1996) 'WW domains of Nedd4 bind to the proline-rich PY motifs in the epithelial Na⁺ channel deleted in Liddle's syndrome.', *The EMBO Journal*, 15(10), pp. 2371–2380. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC450167/>.
- Stephen, A. G. *et al.* (1997) 'Identification of a region within the ubiquitin-activating enzyme required for nuclear targeting and phosphorylation', *Journal of Biological Chemistry*, 272(16), pp. 10895–10903. doi: 10.1074/jbc.272.16.10895.
- Stewart, M. D. *et al.* (2016) 'E2 enzymes: More than just middle men', *Cell Research*, pp. 423–440. doi: 10.1038/cr.2016.35.

- Stone, K. R. *et al.* (1978) 'Isolation of a human prostate carcinoma cell line (DU 145)', *International Journal of Cancer*. Wiley-Blackwell, 21(3), pp. 274–281. doi: 10.1002/ijc.2910210305.
- Sudol, M. *et al.* (1995) 'Characterization of a novel protein-binding module--the WW domain.', *FEBS letters*, 369(1), pp. 67–71. doi: 10.1016/0014-5793(95)00550-S.
- Sutton, R. B. *et al.* (1995) 'Structure of the First C2 Domain of Synaptotagmin 1: A Novel Ca²⁺/Phospholipid-Binding Fold', *Cell*, 80(6), pp. 929–938. doi: 10.1016/0092-8674(95)90296-1.
- Taylor, B. S. *et al.* (2010) 'Integrative Genomic Profiling of Human Prostate Cancer', *Cancer Cell*, 18(1), pp. 11–22. doi: <https://doi.org/10.1016/j.ccr.2010.05.026>.
- Thrower, J. S. (2000) 'Recognition of the polyubiquitin proteolytic signal', *The EMBO Journal*, 19(1), pp. 94–102. doi: 10.1093/emboj/19.1.94.
- Tolchard, J. (2013) 'Insights into the structures and dynamics of the pathogen secreted effectors AVR3A11 and TARP through the application of NMR spectroscopy', PhD thesis, University of East Anglia, UK.
- Trimpert, C. *et al.* (2017) 'NDFIP allows NEDD4/NEDD4L-induced AQP2 ubiquitination and degradation', *PLoS ONE*. Edited by J. L. Brodsky. San Francisco, CA USA: Public Library of Science, 12(9), p. e0183774. doi: 10.1371/journal.pone.0183774.
- Trotman, L. C. *et al.* (2007) 'Ubiquitination Regulates PTEN Nuclear Import and Tumor Suppression', *Cell*, 128(1), pp. 141–156. doi: 10.1016/j.cell.2006.11.040.
- Tsai, J. H. and Yang, J. (2013) 'Epithelial–mesenchymal plasticity in carcinoma metastasis', *Genes & Development*. Cold Spring Harbor Laboratory Press, 27(20), pp. 2192–2206. doi: 10.1101/gad.225334.113.
- Vara, J. Á. F. *et al.* (2004) 'PI3K/Akt signalling pathway and cancer', *Cancer Treatment Reviews*, 30(2), pp. 193–204. doi: <https://doi.org/10.1016/j.ctrv.2003.07.007>. Vassilev, L. T. *et al.* (2004) 'In Vivo Activation of the p53 Pathway by Small-Molecule Antagonists of MDM2', *Science*, 303(5659), p. 844 LP-848. Available at: <http://science.sciencemag.org/content/303/5659/844.abstract>.
- Vazquez, F. *et al.* (2000) 'Phosphorylation of the PTEN Tail Regulates Protein Stability and Function', *Molecular and Cellular Biology*, 20(14), pp. 5010–5018. doi: 10.1128/MCB.20.14.5010-5018.2000.
- Venuprasad, K. *et al.* (2015) 'Multifaceted role of the ubiquitin ligase Itch in immune regulation', *Immunology and Cell Biology*. doi: 10.1038/icb.2014.118.
- Verdecia, M. A. *et al.* (2003) 'Conformational flexibility underlies ubiquitin ligation mediated by the WWP1 HECT domain E3 ligase', *Molecular Cell*, 11(1), pp. 249–259. doi: 10.1016/S1097-2765(02)00774-8.
- Vranken, W. F. *et al.* (2005) 'The CCPN data model for NMR spectroscopy: Development of a software pipeline', *Proteins: Structure, Function and Genetics*, 59(4), pp. 687–696. doi: 10.1002/prot.20449.
- Wahl, L. (2016) 'Decoding the structure and function of WWP2 in the TGFβ signalling pathway', PhD thesis, University of East Anglia, UK.

- Walker, L. *et al.* (2010) 'PDB: 3L4H', Structural Genomics Consortium (SGC), the structure of the helical box domain and second WW domains of the human E3 ubiquitin-protein ligase HECW1 expressed in *E.coli*.
- Walker, L. *et al.* (2013) 'Expression of TGF β 3 and its effects on migratory and invasive behavior of prostate cancer cells: involvement of PI3-kinase/AKT signaling pathway', *Clinical & experimental metastasis*, 30(1), pp. 13–23. doi: 10.1007/s10585-012-9494-0.
- Walker, S. M. *et al.* (2004) 'The tumour-suppressor function of PTEN requires an N-terminal lipid-binding motif.', *Biochemical Journal*, 379(Pt 2), pp. 301–307. doi: 10.1042/BJ20031839.
- Wan, Y. *et al.* (2010) 'Activation of cAMP-Responsive-Element-Binding Protein by PI3 Kinase and p38 MAPK Is Essential for Elevated Expression of Transforming Growth Factor β 2 in Cancer Cells', *Journal of Interferon & Cytokine Research*. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA: Mary Ann Liebert, Inc., 30(9), pp. 677–681. doi: 10.1089/jir.2009.0117.
- Wang, G. *et al.* (2009) 'Transforming Growth Factor- β -inducible Phosphorylation of Smad3', *The Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 284(15), pp. 9663–9673. doi: 10.1074/jbc.M809281200.
- Wang, J. *et al.* (2010) 'Calcium activates Nedd4 E3 ubiquitin ligases by releasing the C2 domain-mediated auto-inhibition', *Journal of Biological Chemistry*, 285(16), pp. 12279–12288. doi: 10.1074/jbc.M109.086405.
- Wang, S. I., Parsons, R. and Iltmann, M. (1998) 'Homozygous deletion of the PTEN tumor suppressor gene in a subset of prostate adenocarcinomas.', *Clinical Cancer Research*, 4(3), p. 811 LP-815. Available at: <http://clincancerres.aacrjournals.org/content/4/3/811.abstract>.
- Wang, S. *et al.* (2003) 'Prostate-specific deletion of the murine Pten tumor suppressor gene leads to metastatic prostate cancer', *Cancer Cell*, 4(3), pp. 209–221. doi: [https://doi.org/10.1016/S1535-6108\(03\)00215-0](https://doi.org/10.1016/S1535-6108(03)00215-0).
- Wang, X. *et al.* (2007) 'NEDD4-1 Is a Proto-Oncogenic Ubiquitin Ligase for PTEN', *Cell*, 128(1), pp. 129–139. doi: 10.1016/j.cell.2006.11.039.
- Watt, J. *et al.* (2018) 'Discovery of small molecule WWP2 ubiquitin ligase inhibitors', *Chemistry – A European Journal* (manuscript accepted), doi: <https://doi.org/10.1002/chem.201804169>
- Wauer, T. and Komander, D. (2013) 'Structure of the human Parkin ligase domain in an autoinhibited state', *EMBO Journal*. doi: 10.1038/emboj.2013.125.
- Wenzel, D. M., Stoll, K. E. and Klevit, R. E. (2011) 'E2s: structurally economical and functionally replete', *Biochemical Journal*, 433(1), pp. 31–42. doi: 10.1042/BJ20100985.
- Wiesner, S. *et al.* (2007) 'Autoinhibition of the HECT-Type Ubiquitin Ligase Smurf2 through Its C2 Domain', *Cell*, 130(4), pp. 651–662. doi: 10.1016/j.cell.2007.06.050.
- Wikström, P. *et al.* (1998) 'Transforming growth factor β 1 is associated with angiogenesis, metastasis, and poor clinical outcome in prostate cancer', *The Prostate*. Wiley-Blackwell, 37(1), pp. 19–29. doi: 10.1002/(SICI)1097-0045(19980915)37:1<19::AID-PROS4>3.0.CO;2-3.

- Willis, M. S. *et al.* (2014) 'The role of ubiquitin ligases in cardiac disease', *Journal of Molecular and Cellular Cardiology*. doi: 10.1016/j.yjmcc.2013.11.008.
- Wu, X. *et al.* (2013) 'Current mouse and cell models in prostate cancer research', *Endocrine-Related Cancer*. doi: 10.1530/ERC-12-0285.
- <https://wiki.nci.nih.gov/display/NCIDTPdata/Compound+Sets>, NIH DTP compound libraries (first accessed: Jan 2017)
- <https://www.lgcstandards-atcc.org/> (2016) DU145 (ATCC HTB-81™), available at: <https://www.lgcstandards-atcc.org/Global/Products/6/1/5/F/HTB-81.aspx>
- <https://www.lgcstandards-atcc.org/> (2016) PC-3 (ATCC CRL-1435™), available at: <https://www.lgcstandards-atcc.org/Products/All/CRL-1435.aspx#generalinformation>
- www.nice.org.uk (2014), Prostate cancer: diagnosis and management, available at: <https://www.nice.org.uk/guidance/cg175> (Accessed 3 Aug 2018)
- www.nice.org.uk (2018) Bortezomib, available at: <https://bnf.nice.org.uk/drug/bortezomib.html> (Accessed: 9 July 2018)
- www.nice.org.uk (2018) Carfilzomib, available at: <https://bnf.nice.org.uk/drug/carfilzomib.html> (Accessed: 9 July 2018)
- www.prostatecanceruk.org (2017), Facts and figures, available at: <https://prostatecanceruk.org/prostate-information/about-prostate-cancer> (Accessed: 7 July 2018)
- Xu, G. W. *et al.* (2010) 'The ubiquitin-activating enzyme E1 as a therapeutic target for the treatment of leukemia and multiple myeloma', *Blood*, 115(11), pp. 2251–2259. doi: 10.1182/blood-2009-07-231191.
- Xu, P., Liu, J. and Derynck, R. (2012) 'Post-translational regulation of TGF- β receptor and Smad signaling', *FEBS letters*, 586(14), pp. 1871–1884. doi: 10.1016/j.febslet.2012.05.010.
- Yan, X. *et al.* (2016) 'Smad7 Protein Interacts with Receptor-regulated Smads (R-Smads) to Inhibit Transforming Growth Factor- β (TGF- β)/Smad Signaling', *The Journal of Biological Chemistry*. 11200 Rockville Pike, Suite 302, Rockville, MD 20852-3110, U.S.A.: American Society for Biochemistry and Molecular Biology, 291(1), pp. 382–392. doi: 10.1074/jbc.M115.694281.
- Yang, B. and Kumar, S. (2010) 'Nedd4 and Nedd4-2: Closely related ubiquitin-protein ligases with distinct physiological functions', *Cell Death and Differentiation*, pp. 68–77. doi: 10.1038/cdd.2009.84.
- Yang, Y. *et al.* (2007) 'Inhibitors of ubiquitin-activating enzyme (E1), a new class of potential cancer therapeutics', *Cancer Research*, 67(19), pp. 9472–9481. doi: 10.1158/0008-5472.CAN-07-0568.
- Yang, Y. *et al.* (2005) 'Small molecule inhibitors of HDM2 ubiquitin ligase activity stabilize and activate p53 in cells', *Cancer Cell*, 7(6), pp. 547–559. doi: 10.1016/j.ccr.2005.04.029.
- Yavari, I. and Roberts, J. D. (1978) 'Differential rates of proton exchange for the guanidinium nitrogens of L-arginine determined by natural-abundance nitrogen-15 nuclear magnetic resonance spectroscopy', *Biochemical and Biophysical Research Communications*, 83(2), pp. 635–640. doi: 10.1016/0006-291X(78)91037-9.

- Yim, E.K. *et al.* (2009) 'Rak functions as a tumor suppressor by regulating PTEN protein stability and function', *Cancer Cell*, 15(4), pp. 304-14. doi: 10.1016/j.ccr.2009.02.012.
- Yim, T. (2017) 'Investigating the role of WWP2 ubiquitin ligase isoforms in TGF β -dependent oncogenic signalling', PhD thesis, University of East Anglia, UK.
- Zhang, W. and Sidhu, S. S. (2014) 'Development of inhibitors in the ubiquitination cascade', *FEBS Letters*, pp. 356–367. doi: 10.1016/j.febslet.2013.11.003.
- Zhang, W. *et al.* (2016) 'System-Wide Modulation of HECT E3 Ligases with Selective Ubiquitin Variant Probes', *Molecular Cell*, 62(1), pp. 121–136. doi: 10.1016/j.molcel.2016.02.005.
- Zhao, Y. *et al.* (2015) 'Small-Molecule Inhibitors of the MDM2–p53 Protein–Protein Interaction (MDM2 Inhibitors) in Clinical Trials for Cancer Treatment', *Journal of Medicinal Chemistry*. American Chemical Society, 58(3), pp. 1038–1052. doi: 10.1021/jm501092z.